

1

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

Bert N. Bakker and Yphtach Lelkes

University of Amsterdam

---

<sup>1</sup>We thank John Bullock, Donald Green, Cindy Kam, Martijn Schoonvelde, Kevin Smith and Marco Steenbergen for comments on earlier versions of this draft. Earlier versions of this manuscript were presented at the Quantitative Methods for Social Science seminar at the University of Amsterdam and the Research Lab of the Political Communication and Journalism Group of the Amsterdam School of Communication Research.

## Abstract

Political scientists are increasingly interested in the interplay between personality and politics. This work, by and large, relies on large omnibus surveys like the ANES or the CCES, where the space and time needed to administer long personality measures is unavailable. We investigate whether the reliance on short personality scales has affected our understanding of personality and politics in two studies. First, we explore whether a popular personality construct (Need for Cognition) moderates the effect of policy cues on attitudes. In line with most extant political science research, we find that NfC does not impact cue taking when we use a 2-item measure of Need for Cognition (NfC). When we use larger measures, we find a substantive and significant effect, which is in line with theories of information processing. Next, we investigate whether the Big Five personality traits are related to various dimensions of ideology. More elaborate measures of personality yield relationships that are about twice as large as those found when we use the ubiquitous 10 item Big Five Inventory (BFI). Traits that been largely dismissed as relevant to political ideology yield stronger and more consistent associations when larger batteries are employed versus the brief BFI. Both studies imply that the measurement of personality conditions the conclusions we draw about the role of personality in politics

**Keywords:** Political ideology, Big Five, Need for Cognition, Measurement

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

The study of personality and politics has – after lying relatively dormant for several decades (but see, McClosky, 1958; Sniderman, 1975) – received renewed interest from political scientists. Work by, among others, Bullock (2011), Gerber, Huber, Doherty, Dowling, and Ha (2010), Hibbing, Smith, and Alford (2014), Johnston (2013) Kam (2012), Kam and Estes (2016), Leeper (2014), Lyons, Sokhey, McClurg, and Seib (2016), Mondak and Halperin (2008) and Petersen and Aarøe (2013), has led to new and important insights into the relationship between personality traits and political attitudes and behaviors. By and large, political scientists have adopted a different methodological approach to the study than that of most psychologists, which, we contend, has yielded sometimes inaccurate results.

While psychologists tend to rely on convenience samples, most of the extant knowledge on public opinion and political behavior is based on cross-sectional and longitudinal omnibus surveys with excellent probability samples such as the American National Elections Studies and the British National Election Studies. The trade-off for the higher quality samples is shorter measures of the constructs we are interested in. The 2012 post-election ANES, for instance, contained hundreds of questions which purportedly measure over a 150 broadly defined constructs. While textbooks on measurement recommend long scales over short scales (Cronbach, 1949), political science tends to ignore this advice and latent constructs are often measured using one or two items.

Widely used instruments used to measure personality in political science are radically shorter than those originally developed in psychology. The personality batteries are often culled from fairly long and often multi-dimensional scales developed and validated by psychologists. The often used Big Five Inventory (Henceforth, BFI) and Ten Item Personality Inventory (henceforth, TIPI) measuring the Big Five personality traits using 10 items is between five and 24 times shorter than other measures of the same construct. As is well known, such short forms are less reliable and

may measure only some of the lower-order sub-dimension of the trait. As a consequence, it is likely that studies relying upon brief measures of personality suffer from both Type M (magnitude) and Type S (sign) errors compared to the more elaborate measures (Gelman & Carlin, 2014). Despite these risks, short personality measures continue to be mainstay of personality and politics research in political science.

In this paper, we assess the impact of the trend in political science to use short measures of personality. We focus on two central debates within the literature. The first touches upon the role of the role of Need for Cognition (NfC) – the tendency to enjoy thinking (Cacioppo & Petty, 1982) – in moderating the reliance upon 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 both studies, we show that the brief measures which are fairly ubiquitous in the political science literature would lead us to Type M and Type S errors. We also find that only a slight increase in the number of items we use yields outcomes consistent with the more elaborate measures.

Our paper has important substantive implications. First, the majority of political science studies have relied upon an abbreviated 2-item NfC measure (see for instance, Kam, 2005) and do not find evidence consistent with the Elaboration Likelihood Model. We show that this conclusion is an artifact of the measure – using a full 18-item battery, NfC moderates the reliance upon policy information. Second, we show that, if we rely upon the commonly used BFI, then we conclude that traits such as Agreeableness, Extraversion and even Conscientiousness as irrelevant for politics, while other traits are weakly associated with politics. Yet, once we rely upon a more elaborate battery, many of the Big Five personality are as highly correlated with the same political outcomes as the traits commonly shown to be important for politics, namely Openness and Conscientiousness (Jost, Glaser, Kruglanski, & Sulloway, 2003).

This work also has important methodological implications. Brief measures of personality lead to Type M and in some instances even to Type S errors compared to larger measures. However, we show that we need not saturate our questionnaires with

extremely long batteries. Rather, we put forward and confirm that increasing short batteries only slightly – adding 2-4 items per trait – yields, results that alleviate Type S and Type M errors. We also argue that criterion validity should receive more attention in political science research which has predominantly focused upon assessment of reliability, over time stability and content validity of (psychological) constructs. Related, the ongoing debate on replicability should also consider the measurement of constructs, in addition to important issues such as accessibility of data (Lupia & Elman, 2014), replication and pre-registration (Mullinix, Leeper, Druckman, & Freese, 2015), data-quality (Berinsky, Huber, & Lenz, 2012) and null-hypothesis testing (Gill, 1999), in the ongoing debate about the *good* research practices in political science research.

### **The consequences of using brief personality measures**

Brief measures of personality offer several advantages over their longer counterparts. Short surveys are cheaper to administer (Smits & Vorst, 2007) and increase the response rate (Edwards, Roberts, Sandercock, & Frost, 2004). The boredom and fatigue that may arise as a result of completing a long index may increase measurement error (Burisch, 1984a). 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 on a survey. Fortunately, some of the psychometric properties of the BFI, the NfC as well as other brief personality measures are satisfactory. The test-retest reliability of the BFI (Rammstedt & John, 2007), TIPI (Gosling, Rentfrow, & Swann, 2003) and the NfC (Sadowski & Gulgoz, 1992) are acceptable. Likewise the convergent validity of brief measures is satisfactory. For instance, when ten item measures were compared to the personality batteries with a modest length the convergent validity was acceptable for the BFI (Rammstedt & John, 2007) and the TIPI (Donnellan, Oswald, Baird, & Lucas, 2006; Ehrhart et al., 2009; Furnham, 2008; Gosling et al., 2003; Hofmans, Kuppens, & Allik, 2008; Rojas & Widiger, 2014).

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). Why is the case? Brief measures of personality share common wording, response format and are asked in rapid succession which suggests there is non-random measurement error (Green, 1988; Green, Goldman, & Salovey, 1993). The measurement error in the independent variables attenuates the relationship with certain criterion (i.e., outcome) measures – a process called regression dilution. The correlation between an item and its criterion cannot be larger than the square-root of the reliability of the item (Niemi, Carmines, & McIver, 1986).<sup>2</sup>

Second, personality research is particularly sensitive to short batteries because of the so-called “bandwidth-fidelity” trade-off introduced by Cronbach and Gleser (1957). Bandwidth is “the amount of complexity of the information one tries to obtain in a space of time” (Cronbach, 1949, p.600). Many personality traits are fairly complex constructs with broad bandwidth. For instance, the Five Factor Model of personality suggests that each of the five broad domains – i.e., Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism – consists of six lower order facets. Conscientiousness, for instance, contains the facets Achievement Striving, Competence, Dutifulness, Deliberation, Self-Discipline and Order. Short batteries, such as the BFI and TIPI, may only tap into – at best – a few of these lower order facets. 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, Harms, Niehorster, & Gaye-Valentine, 2012). Conversely, if a target measure is only related to one specific sub-dimension of that trait, but not others, any correlation will be overestimated (Credé et al., 2012; Messick, 1995; Paunonen & Jackson, 1985a). Accordingly, there is the risk of a Type M (magnitude) error (Gelman & Carlin, 2014). Overestimation of relationships should not be a problem for low bandwidth traits, such as those often

---

<sup>2</sup>Niemi et al. (1986) pointed out that adding more items to a scale does not *necessarily* increase the *reliability* nor *validity* of the scale. Specifically, adding an item to a scale that has a low inter-item correlation with the other items does not add a lot to the reliability of the scale and could even decrease the association with a criterion measure so not contributing to the validity. Yet, adding an item with a relatively high inter-item correlation increases the reliability of the scale and the validity of the scale.

considered in the information processing literature (Need for Affect, Need to Evaluate, Need for Cognition, etc.). However, short measures may lead us to incorrectly dismiss the importance of these traits. Finally, and perhaps the worst case scenario, we even would see a Type S (sign) error: this is the case if a specific sub-dimension of a trait is differentially correlated with a criterion measure compared to the broader trait. To summarize, the criterion validity seems to be affected by the use of brief measures (Credé et al., 2012; Smith, McCarthy, & Anderson, 2000).

In this study we assess the consequences of using brief personality measures. We investigate whether the length of the battery used to measure a personality trait conditions the strength and direction of the relationship between personality and politics.<sup>3</sup> Utilizing a unique set of data, wherein respondents were asked multiple forms of the same construct, we test whether we are selling ourselves short when relying upon brief measures of personality. First, we look at whether Need for Cognition (Cacioppo & Petty, 1982) moderates the impact of policy information on political attitudes. Because of the unidimensionality of NfC, we would expect to see regression dilution as measures get shorter.<sup>4</sup> Therefore, we expect to see Type M errors: meaning that we would reject the trait as relevant for politics, while longer measures yield the opposite result. Next, we turn to assessing the role of the Big Five personality traits in political ideology. Big Five traits are complex and thus have a high bandwidth (John, Hampson, & Goldberg, 1991). Therefore we expect both Type M and Type S errors when we rely upon brief measures of personality.

---

<sup>3</sup>According to some the criterion validity of brief scales was acceptable compared to longer measures of different personality scales when focusing upon non-political outcomes (Burisch, 1984b; Paunonen & Jackson, 1985b; Robins, Hendin, & Trzesniewski, 2001; Thalmayer, Saucier, & Eigenhuis, 2011a).

<sup>4</sup>Cronbach and Gleser (1957) would define this as a construct with a relatively high degree of fidelity – i.e., a narrow construct.

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

The Elaboration Likelihood Model, which underpins much of the political persuasion literature (Alvarez & Brehm, 1995; Johnson & Martin, 1998), argues that individual differences in NfC moderates the extent to which citizens' political attitudes are influenced by policy information. NfC captures individual differences in the tendency to enjoy thinking (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984). Those high on NfC tend to 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).

The two-item version of the NfC, which was originally developed for inclusion in the 2000 ANES (Bizer, Krosnick, Petty, Rucker, & Wheeler, 2000), is frequently used in political science research to test for individual differences in the reliance upon policy information. The ANES measure is a highly shortened form of the 18-item NfC scale (Cacioppo et al., 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 generally fail to find evidence that NfC moderates the impact of policy information on political attitudes (Berinsky & Kinder, 2006; Holbrook, 2006; Kam, 2005; Mérola & Hitt, 2015; Rudolph, 2011; Sokhey & McClurg, 2012). Holbrook (2006, 349) acknowledged the possibility that the "weak performance" of the NfC is explained from the fact "that the two survey items used to measure Need for Cognition are adapted from a broader ... measure and may not do an adequate job of representing the underlying concept." Employing a somewhat larger 6-item NfC battery, Bullock (2011) finds that NfC *does* moderate the effect of policy information on policy attitudes. In an exploratory analysis, NfC *does not* moderate the effect of policy cues when Bullock (2011) subsetted the 6-item battery to the 2-item battery employed. Importantly, Bullock (2011) does not compare the 2-item measure to a validated NfC battery (Cacioppo & Petty, 1982; Cacioppo et al., 1984), so we have to be cautious in over-interpreting these results. To determine whether citizens' tendency to rely upon



policy information is moderated by NfC, we conducted the following study.

## Data and Methods

### Design

We analyzed a survey experiment designed by a separate team to test the effect of policy information on policy attitudes (Coffe, 2013). The experiment was fielded on the Longitudinal Internet Studies for the Social sciences (LISS) panel, and consisted of a 2 (center-right vs. radical-right message) X 2 (no party cue vs. party cue) X 2 (aggressive vs. nuanced manner of speech) X 2 (male vs. female politician) experimental design. Participants were shown a professionally edited campaign video in which a randomly assigned to a center-right message or radical-right message. For instance, discussing the issue of immigration, the *center-right* candidate stated, “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* candidate stated, “We demand a complete cessation of immigration of people from Islamic countries.”<sup>5</sup> Each treatment lasted approximately two minutes. Given that our interest lies in the extent to which participants rely upon policy cues (center-right vs. radical-right message), we controlled for variables indicating the other conditions of the experiment, but do not discuss them in detail.

### Participants

We used three waves of the Dutch LISS panel (Scherpenzeel & Das, 2010). The LISS panel is a true probability sample of Dutch households drawn from the official population registry. Panelists who did not have a computer and Internet access were provided one. These panelists complete monthly surveys for about 15 euros per hour. Between 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). Participants were asked to

---

<sup>5</sup>See Appendix X for full text of all experimental treatments.

indicate if they could *hear* and *see* the video. In total 455 (i.e., 8.77%) participants indicated that they could not hear and/or see the video. These participants were excluded from further analyses.

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 2012 (79.3% response rate) and May 2011 (74% response rate). By merging these waves, we had a measure of the Need for Cognition for almost all participants (N=4,553).

## Measures

*Dependent variable.* 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).

*Need for Cognition* was measured using a Dutch version of the 18-item battery developed by Cacioppo et al. (1984). The items of this personality inventory were translated into Dutch by professional translators, using the widely accepted translation-back-translation method. The principle investigators of the panel resolved inconsistencies in the translations. The 18-item 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). Table 1 provides the descriptive statistics of the three NfC measures and indicates that three NfC batteries are strongly correlated.

## Results

Figure 1 presents the results of the regression models and plots the fitted estimates of agreement with the center-right and radical-right message over the range of the Need for Cognition, while controlling for the other conditions in the experiment. Using the two item ANES measure, we find that NfC *does not* moderate the effect of policy messages on policy attitudes ( $b=-0.07$ ,  $SE=0.04$ ). That is, as NfC increases,

Table 1

*Descriptive statistics Need for Cognition*

Battery	Mean	SD	Skew	Kurtosis	$\alpha$	$r_{\text{Bullock}}$	$r_{\text{ANES}}$
full 18-item	0.55	0.15	-0.13	3.05	.90	.91	.78
Bullock 6-item	0.55	0.18	-0.09	2.81	.77	1	.83
ANES 2-item	0.54	0.24	-0.32	2.55	.84	.83	1

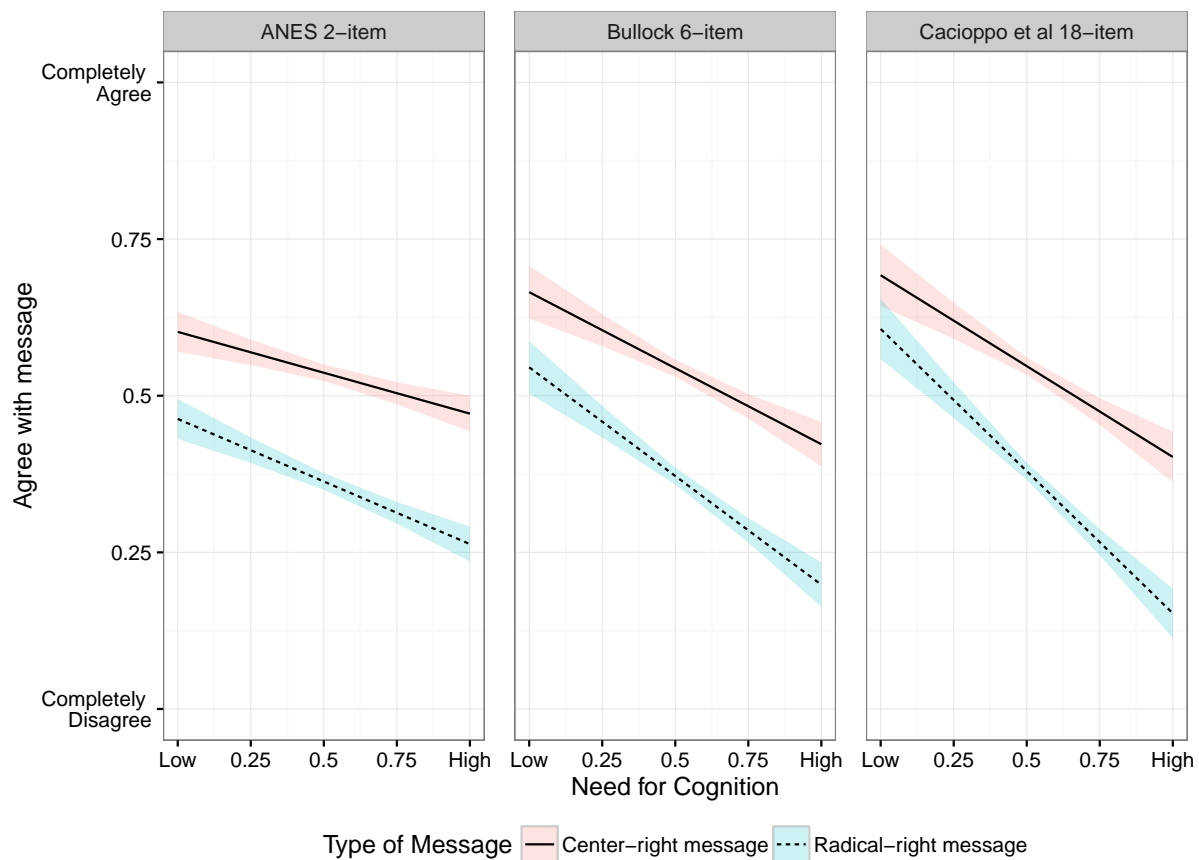
*Note: N = 4,733*

agreement with the policy decreases irrespective of the message, which is consistent with the finding that NfC is negatively correlated with right-wing ideology (Ksiazkiewicz, Ludeke, & Krueger, 2016). Most importantly, relying upon the ANES measure of the NfC would yield additional support for the body of political science research which rejects NfC as a moderator of policy cues.

When we rely upon more elaborate measures of the NfC, we arrive at conclusions more in line with the Elaboration Likelihood Model. When running a similar model using the 18-item NfC battery we find a significant interaction effect that is more than twice the size of the ANES-based estimate of the interaction effect ( $b=-0.16$ ,  $SE=0.06$ ). Specifically, the difference in the agreement between the right-wing and radical-right messages increases as NfC goes up (see middle panel of Fig 1). Moreover, we subsetted our 18-item measure to the 6-item measure employed by Bullock (2011). Here, the interaction is significant ( $b=-0.10$ ,  $SE=0.05$ ).

One possibility is that our results are 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 the 18 items, we generated all possible combinations for scales of different length. This resulted, for instance, in 18 different 1-item measures, 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 262,145 instantiations of NfC, we then

Figure 1. Need for Cognition and Policy Cues

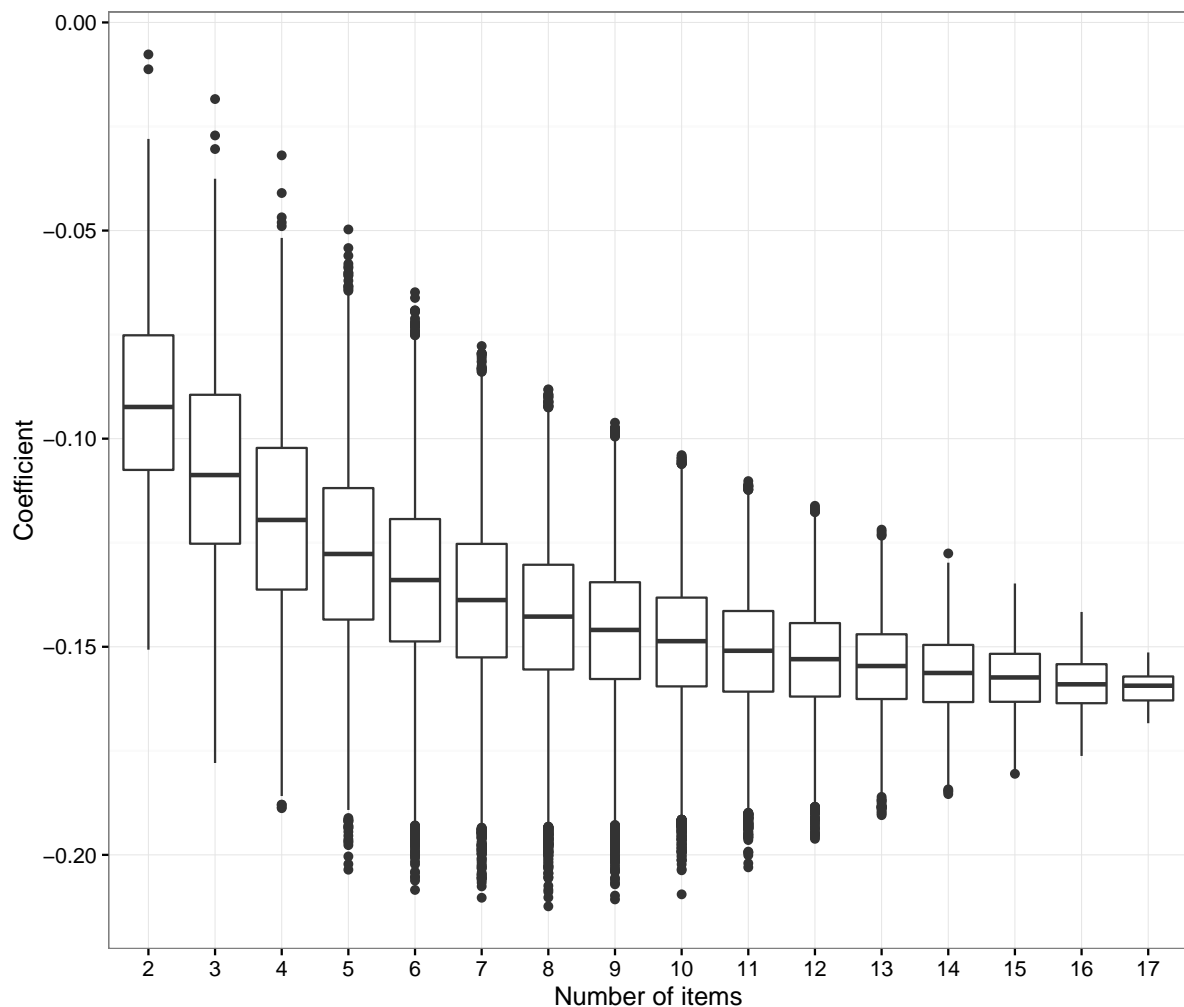


*Note: OLS estimates with 95% Confidence Intervals are plotted. The Supplementary Material provides the tables with results of the regression models for the models including the ANES 2-item (Table X), Bullock 6-item (Table X) and 18-item measure (Table X).*

calculated the interaction effects between receiving a center-right or radical-right message 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.

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

Figure 2. Need for Cognition and Policy Cues



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.

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). We show, in Supplementary Material X, 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.

Because Cronbach's Alpha increases with the number of items in the scale, we replicate these results with McDonald's Omega (Ziegler, Kemper, & Krueger, 2014), another measure of reliability that is less dependent upon the length of the scale (McDonald, 1970).<sup>6</sup> Yet, we show in the Supplementary Materials that a brief scale with a higher Omega coefficient does not lead to more reliable estimates compared to a brief scale with a lower Omega coefficient. Why do Cronbach's Alpha and McDonald's Omega not result in satisfactory selection of items? A possible answer is the fact that these measures of reliability assume random measurement error. The NfC - as well as other personality items - share common wording, response format and are asked in rapid succession which suggests there are "response set" errors (Green, 1988; Green et al., 1993).

Finally, common practice in selecting items to form a short measure is to select items that load most highly in a confirmatory factor analysis of the items in the larger method. This was the method used to develop the 2-item NfC ANES measure, for instance, as well as the Bullock 6-item measure. 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. This suggests that relying solely upon Cronbach's alpha, high factor item loadings or McDonald's Omega do not result in brief measures of personality with a high criterion validity.

The Need for Cognition 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 our errors were namely one of magnitude (i.e., Type M, Gelman & Carlin, 2014). Some personality measures are much more complex. What are consequences of using short measures for high bandwidth measures?

---

<sup>6</sup>Omega can only be calculated when the scale contains three items or more so we cannot assess the 2-item measure using Omega.

### Short measures with high bandwidth: The Case of the BFI

Traditionally, high bandwidth personality traits 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. 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 as well as the British National Election Studies.

The Big Five personality traits – Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism – are high bandwidth constructs. The items for a short measure of each trait are selected so that the short measure reflects the breadth of the original dimension. Accordingly, the inter-item correlation will be low (Woods & Hampson, 2005; ?), necessarily decreasing the reliability and criterion validity of the scale. Second, and more fundamentally, all personality dimensions should capture variance in broad Big Five traits (Credé et al., 2012; Smith et al., 2000). However, some aspects of a trait will be underrepresented in a short measure, which limits the content validity of the trait.

The burgeoning literature on the association between personality – especially the so called Big Five traits – and political ideology (e.g., Carney, Jost, Gosling, & Potter, 2008; Gerber et al., 2010; Mondak & Halperin, 2008) could be particularly prone to the detrimental consequences of using brief measures. Brief measures of personality – in which each trait is measured with 1, 2 or 3 items – have been used extensively in recent years, to study the relationship between personality and political ideology (see for instance, Ackermann & Ackermann, 2015; Ackermann, Ackermann, & Freitag, 2016; Bakker, 2016; Carney et al., 2008; Clifford, Jewell, & Waggoner, 2015; Dietrich, Lasley, Mondak, Rempel, & Turner, 2012; Dietrich et al., 2012; Fatke, 2016; Freitag & Rapp, 2015a, 2015b; Friesen & Hibbing, 2016; Gerber, Huber, Doherty, & Dowling, 2011; Gerber et al., 2010; Ha, Kim, & Jo, 2013; Jost, West, & Gosling, 2009; Nielsen, 2016; Rentfrow et al., 2013; Rentfrow, Jost, Gosling, & Potter, 2009; Sweetser, 2013,?;

Weinschenk, 2014). While many of these studies have yielded fairly consistent results when it comes to direction and statistical significance (although not strength), for a number of traits, different studies yield different conclusions.

For instance, Openness, regardless of measurement, has yielded consistent correlations with unidimensional conservatism (Sibley, Osborne, & Duckitt, 2012) and social conservatism (Carney et al., 2008; Gerber, Huber, Doherty, & Dowling, 2011; Gerber et al., 2010; Leeson & Heaven, 1999; Riemann, Grubich, Hempel, Mergl, & Richter, 1993). Scholars have also consistently found no relationship between Openness and economic conservatism (Bakker, 2016; Carney et al., 2008; Fatke, 2016; Van Hiel & Mervielde, 2004). On the other hand, many of the results of studies looking at Neuroticism conflict. Using elaborate measures, some have found no relationship between Neuroticism and cultural ideology (Bakker, 2016; Leeson & Heaven, 1999; Riemann et al., 1993; Van Hiel & Mervielde, 2004; Verhulst, Eaves, & Hatemi, 2012; Verhulst, Hatemi, & Martin, 2010), while others, using brief measures found positive associations (Gerber, Huber, Doherty, & Dowling, 2011; Gerber et al., 2010). Similarly, there is some evidence that Neuroticism correlates negatively with economic conservatism (Bakker, 2016; Gerber, Huber, Doherty, & Dowling, 2011; Gerber et al., 2010; Verhulst et al., 2012), but other studies that other studies do not report these associations (Carney et al., 2008; Leeson & Heaven, 1999; Van Hiel & Mervielde, 2004). Could some of the inconsistencies in the personality and politics literature be explained by the use of brief personality scales? And, as a consequence of the null findings of many past studies, are we incorrectly dismissing the relevance of certain personality characteristics for politics?

In order to answer these questions, one would need to compare the association between the traits from different personality inventories and political outcomes. As far as we know, there have been only two preliminary assessments of the consequences of scale length on the association between personality and politics. Mondak, Hibbing, Canache, Seligson, and Anderson (2010, Appendix A) found that – using a self-constructed Big Five battery – Openness and Conscientiousness correlated



consistently with a uni-dimensional measure of ideology when each trait was measured using two or five items per trait. However, the results for the other three traits were less consistent across the two batteries. An “exploratory analysis” by Gerber, Huber, Doherty, and Dowling (2011, 280) 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), which is surprising—since the TIPI is less reliable, we would expect correlations to be generally weaker. This, “suggests researchers should be sensitive to the consequences of using different personality batteries for predicting political outcomes” (Gerber, Huber, Doherty, & Dowling, 2011, 280). These necessarily preliminary studies did not offer an in-depth exploration of *why* measurement of the personality trait would condition the criterion validity. We extend the current state-of-the-art by pointing out that (1) regression dilution and (2) the bandwidth of the trait could lead to both Type M and Type S errors.

We compare the criterion validity of the BFI with the 50-item Five Factor Model International Personality Item Pool (IPIP; Goldberg, 1999). The IPIP-FFM has a couple of advantages. First, the battery intends to measure the NEO PI-R – the most well known battery of personality (IPIP-FFM Donnellan et al., 2006; Goldberg et al., 2006; Gow, Whiteman, Pattie, & Deary, 2005). Yet, the IPIP-FFM is publicly available – compared to the NEO PI R which is a proprietary battery which is rarely employed in the study of politics – and has recently been employed in the study of personality and politics (Bakker, Rooduijn, & Schumacher, 2016; Gallego & Pardos-Prado, 2014). The IPIP-FFM shows good convergent validity with the 10-item BFI (Thalmayer, Saucier, & Eigenhuis, 2011b), the 44-item BFI (Donnellan et al., 2006) and the TIPI (Ehrhart et al., 2009).<sup>7</sup> Another advantage is that using the 50-item IPIP one can also derive the Mini-IPIP, a validated and reliable 20-item battery (i.e., 4 items per trait Donnellan et al., 2006), which has recently also been used in the study of personality and politics

---

<sup>7</sup>Silvia and Sanders (2010) showed that the Openness dimension of the IPIP FFM, TIPI and BFI correlate  $>.5$

(Bakker & de Vreese, 2016; De Neve, 2015; Osborne & Sibley, 2012, 2015). Accordingly, this set-up allows us to compare the association between personality and ideology using the BFI to validated and reliable 4-item (i.e. Mini-IPIP) and 10-item per trait (i.e. IPIP-FFM) measures.

## Data and Methods

### Participants

We used two waves of the Dutch Longitudinal Internet Studies for Social sciences (LISS) panel (Scherpenzeel & Das, 2010). In July of 2012, LISS panelists (N=5,708) were invited to complete a survey that included a 50-item personality inventory. Five months later, a subset of these respondents took part in the Dutch module of the 2012 World Values Survey.<sup>8</sup> The response rate was 76.6% (N=1,901) and the completion rate was 76.0% (N=1,884). Our sample was restricted to those participants who filled out both waves, this results in a data-set with 1,582 respondents.<sup>9</sup>

### Measuring personality

Personality was assessed using two different batteries, in two waves of the LISS panel. Participants completed a 50-item personality inventory in the first wave, namely the International Personality Item Pool – Five Factor Model (IPIP-FFM) (Goldberg et al., 2006). The items of the IPIP-FFM 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. A unique aspect of the 50-item IPIP-FFM is that it also possible to derive a validated and reliable 20-item instrument, the Mini-IPIP (Donnellan et al., 2006). Accordingly, we could assess the relationship between personality and political ideology using personality inventories that measure each trait using 2-items (i.e., BFI), 4-items (i.e., Mini-IPIP) and 10-items (i.e., IPIP-FFM) per trait. The BFI was administered as part of the WVS

---

<sup>8</sup>This means the survey was conducted in Dutch.

<sup>9</sup>This means we have a data-set that includes 84% of the respondents that completed the WVS 2012 in the Netherlands.

2012 in the Netherlands.

The BFI is thereby measured at the same time as the criterion measures, while the IPIP measures was measured *prior* to the criterion measures. We do not think that this is problematic. Since personality traits are relatively stable over shorter time periods (Gerber, Huber, Doherty, & Dowling, 2013) – and are stably associated with political attitudes over time (Bloeser, Canache, Mitchell, Mondak, & Poore, 2015) – the lag between the two waves should not affect the nature of the associations reported here. Moreover, if anything, we should bias the strength of the associations between personality and the criterion measures in favor of the BFI compared to the measures collected earlier.

For each of the personality inventories employed in this study, we created additive scales of each of the Big Five traits. Each measure was recoded to range from 0 (lowest observed value on the trait) through 1 (highest observed value on the trait). Summary statistics and Cronbach's  $\alpha$  for each trait of the three batteries of personality appear in Table 2. As expected – given the low number of items per trait – BFI traits are less reliable than either the Mini-IPIP or IPIP-FFM. Agreeableness, and to a lesser extent Openness, Conscientiousness, Extraversion and Neuroticism suffered from low internal consistencies.<sup>10</sup> Clearly, the 20-item Mini-IPIP and the 50-item IPIP-FFM did not suffer from this problem and had acceptable internal consistencies. While the standard deviations tended to be slightly larger among the BFI traits than the Mini-IPIP and IPIP-FFM, the means were very similar. The correlations between the BFI and the IPIP measures ranged from weak (Agreeableness) to modestly strong (Neuroticism) (see also, Ehrhart et al., 2009; Goldberg et al., 2006). Finally, we note that the high correlations between the IPIP and Mini-IPIP traits are expected as the Mini-IPIP is created by subsetting the items of the 50-item IPIP.

---

<sup>10</sup>The internal consistency of the BFI is comparable to other studies using population based samples.

Table 2

*Descriptive statistics*

Trait	Battery	Mean	SD	Skew	Kurtosis	$\alpha$	$r_{\text{BFI}}$	$r_{\text{Mini-IPIP}}$
Openness	BFI	0.54	0.25	-0.08	2.34	0.46	-	
	Mini-IPIP	0.59	0.18	0.04	3.09	0.60	0.47	-
	IPIP-FFM	0.57	0.14	0.07	2.95	0.76	0.45	0.84
Conscientiousness	BFI	0.73	0.21	-0.56	2.66	0.42	-	
	Mini-IPIP	0.67	0.18	-0.39	2.83	0.67	0.46	-
	IPIP-FFM	0.68	0.13	-0.29	2.93	0.77	0.51	0.89
Extraversion	BFI	0.53	0.23	0.19	2.64	0.41	-	
	Mini-IPIP	0.55	0.19	-0.07	2.84	0.75	0.46	-
	IPIP-FFM	0.54	0.17	-0.01	2.92	0.87	0.51	0.93
Agreeableness	BFI	0.69	0.18	-0.40	3.09	0.19	-	
	Mini-IPIP	0.72	0.16	-0.60	3.66	0.76	0.22	-
	IPIP-FFM	0.68	0.15	-0.51	3.81	0.83	0.24	0.91
Neuroticism	BFI	0.35	0.24	0.42	2.61	0.68	-	
	Mini-IPIP	0.34	0.19	0.39	3.01	0.73	0.62	-
	IPIP-FFM	0.37	0.17	0.38	2.91	0.88	0.67	0.93

*Note:*  $N = 1582$

### Criterion Measures

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 the primary focus of the personality-politics research. Specifically, we focus upon a uni-dimensional operationalization of ideology (Jost et al., 2003; Mondak & Halperin, 2008) as well as economic and cultural ideology (Feldman & Johnston, 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 ( $M = 0.50$ ,  $SD = 0.23$ ).

*Cultural ideology* was part of the WVS 2012 and measured using two items, namely “I find it shocking when a man and a woman kiss in public” and “I find it shocking if two men kiss in public.” Both items were scored on a five point Likert-scale ranging from “strongly disagree” through “strongly agree”. The items correlate highly ( $r = 0.62$ ) and were internally consistent ( $\alpha = 0.75$ ). Accordingly, we created a scale ranging from the most liberal (0) through the most conservative (1) cultural ideology ( $M = 0.38$ ;  $SD = 0.24$ ).

*Economic ideology* was also measured using two items in the WVS 2012. The first item asked people to rate themselves on a scale from (1) “Government should take more responsibility to ensure that everyone is provided for” through (10) “People should take more responsibility to provide for themselves.” The second item asked participants to rate themselves on a scale range from (1) “Incomes should be made more equal” through (10) “Individual effort should be rewarded.” We created a scale ( $\alpha = 0.76$ ) ranging from the most liberal (0) through the most conservative (1) economic ideology ( $M = 0.52$ ,  $SD = 0.21$ ).

The three ideology dimensions are conceptually distinct. Unidimensional ideology correlates weak with with cultural ideology ( $r=0.18$ ) and modestly strong with economic ideology ( $r=0.53$ ). Cultural and economic ideology are also weakly associated with each other ( $r=0.07$ ).

## Results

For each personality battery, we regressed – using OLS regression models – the criterion measures on each trait as well as sex, age, education and income. The descriptive statistics of the control variables are provided in Table X of the *Supplementary Material*.<sup>11</sup> To make the results easily comparable, we plot regression coefficients and 95 percent confidence intervals in one figure with a row for each trait

---

<sup>11</sup>The results do not change if we do not control for education and income.

and a column for the associations between each ideology dimension and the BFI, Mini-IPIP and IPIP. We discuss the results trait by trait. The *Supplementary Material* provides the tables with results of the regression models for the models including the BFI (Table X), the Mini-IPIP (Table X) and the IPIP-FFM (Table X). We discuss the results for the three ideology measures on a trait by trait basis.

The number of items used to measure Openness did not substantively change its relationship to almost all the criterion measures (Figure 3). Higher levels of Openness were negatively correlated with conservatism (column 1) and cultural conservatism (column 2). 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 a small positive relationship (but overlapping with zero) between economic ideology and Openness (see Figure 3, row 1, column 3). Overall, this confirms the generally held observation that Openness is the most reliable correlate of political ideology (Sibley et al., 2012) as well as cultural ideology (Carney et al., 2008). The non-significant associations between Openness and economic ideology confirm the suggestion by Hibbing et al. (2014, p. 304) that Openness is “less relevant to economic issues.”

The relationship between Conscientiousness and the various criterion measures was conditional upon the measurement of personality (Figure 3, row 2). There was a consistent association between Conscientiousness and the unidimensional measure of conservatism (Figure 3, row 2, column 1). However, the association between Conscientiousness and conservatism was almost two times larger when measured using the IPIP compared to the BFI with a marginally significant difference in the size of the estimate of 0.13 (SE=0.07,  $p < 0.1$ , two-tailed). Turning to the relationship between Conscientiousness and cultural conservatism and economic conservatism (Figure 3, row 2, column 2 & 3), we would likely conclude that there was no relationship between Conscientiousness and these ideology dimensions because the 95 percent confidence intervals of the BFI and the Mini-IPIP contain zero. Using the more elaborate IPIP, Conscientiousness and cultural and economic conservatism were positive and

significantly correlated. This difference is especially profound for the association between Conscientiousness and cultural conservatism, which was twice as large – and statistically significant ( $b=0.16$   $SE=0.07$ ,  $p<0.05$ ) – when we used the IPIP compared to the BFI. Columns 2 and 3 of Figure ??, row 2, indicate that scholars using the underestimate the size of the association between Conscientiousness and ideology by half, compared to the IPIP. Again these results suggest that we are prone to make Type M errors when we rely upon brief measures of personality.

Unidimensional ideology and Neuroticism were consistently unrelated to each other (Figure ??, row 3, column 1) but the relationship between Neuroticism and the other measures of ideology were conditional upon the operationalization of personality (see row 3, column 2 & 3). Neuroticism was consistently related to cultural conservatism but the size of the association was twice as large when measured with the IPIP compared to the BFI ( $b=0.10$ ,  $SE=0.05$ ,  $p<0.05$ ). The BFI seemed to overestimate the association between Neuroticism and economic conservatism. Specifically, we would conclude that there was a negative correlation between Neuroticism and economic conservatism based upon the BFI (Gerber et al., 2010). Yet, estimates using the more elaborate IPIP indicate that Neuroticism was unrelated to economic conservatism. This result suggests that we are likely to make a Type M – and conclude the trait is relevant for politics – when studying the association with economic ideology.

Measurement conditions the substantive conclusions we draw about the association between Agreeableness and the dimensions of ideology (see Figure 3, row 4). The negative association between Agreeableness and conservatism (row 4, column 1) is three times as large – and statistical significant ( $b=-0.16$ ,  $SE=0.05$ ,  $p<0.05$ ) – 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 *ten* times larger – and statistical significant ( $b=-0.22$ ,  $SE=0.07$ ,  $p<0.05$ ) – compared to the IPIP. Similarly the Mini-IPIP estimate was roughly *seven*

times larger – and statistically significant ( $b=-0.18$ ,  $SE=0.06$ ,  $p<0.05$ ) – than the BFI estimate. 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 twice as large and significant when we use the IPIP. The results for Agreeableness consistently suggest that we underestimate the association between Agreeableness and ideology relying upon the BFI and thus commit a Type M error.

Extraversion yielded the most striking results with both Type M and Type S errors occurring. Using the BFI, the association between Extraversion and unidimensional ideology was negative and not different from zero ( $b=-0.04$ ,  $SE=0.03$ ) (see also [Carney et al., 2008](#); [Gerber, Huber, Doherty, & Dowling, 2011](#); [Sibley et al., 2012](#)). The IPIP estimate was positive and five times larger and statistically significantly stronger ( $b=0.20$ ,  $SE=0.06$ ,  $p<0.05$ ) compared to the BFI estimate. The Mini-IPIP estimate was four times larger than the BFI estimate. While those using the BFI would conclude that the relationship between cultural ideology and Extraversion was negative, those using the IPIP would probably argue that there is no relationship between the two constructs. Those using the BFI would find no relationship between Extraversion and economic ideology, while those using either the Mini-IPIP or the IPIP would find a positive one.

One may say that our results are due to the idiosyncrasies of the TIPI. Perhaps another brief personality inventory – such as the Five Item Personality Inventory (?), Single Item Measure of Personality ([Woods & Hampson, 2005](#)), Ten Item Personality Inventory ([Gosling et al., 2003](#)) – would result in estimates more consistent with larger personality batteries. Unfortunately, our study does not contain any of these alternative measures. We can, however, using the set-up of our study address the issue using a different approach. For each set of the 10 items that theoretically make-up a trait, we generated 10 1-item measures, 45 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 then 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. Figure 4 plots the distribution of the point estimates of these measures sorted by the number of items used to generate the trait. We plot distribution of point estimates of the regression coefficients from models predicting each form of ideology (the columns) to each 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.

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. Take, for instance, the fourth row of Figure 4, which displays the regression coefficients from models predicting each form of Agreeableness from ideology. Starting with the left most column, where the outcome measure is unidimensional ideology, the median association between Agreeableness made up of one item from the ten item scale is roughly  $-.07$ . The median association more or less linearly increases as we increase the number of items – the median association between a three item scale is roughly  $-.12$ ; the median association between a six item Neuroticism scale is roughly  $-.16$ ; and the association between a 9-item scale is roughly the same as that of the IPIP. Note as well that the estimate generated from the Mini-IPIP is almost exactly the same as the median estimate from randomly chosen 4-item scales. The association from generated from the BFI's Agreeableness scale was roughly the same as the median estimate from 1-item measures. It seems that one would do better if they randomly chose two items to make up an Agreeableness scale than if they used the BFI.

A similar pattern emerges in the next panel; plotting the point estimates from a model regressing cultural ideology and Agreeableness. The median association between 1-item measures and the outcome measure is roughly  $-.13$ ; the relationship roughly doubles at a fairly linear rate as we increase to the full 10-item IPIP Agreeableness measure. The BFI Agreeableness measure was much smaller than almost any randomly selected two-item measures and all but one one item measures. The relationship between Neuroticism and unidimensional and cultural ideology are also contingent on

the number of items used to make these scales.

### Discussion

This paper makes both a series of substantive and methodological contributions. Counter to many past findings – which have utilized the 2-item ANES measure –, but in line with the Elaboration Likelihood Model of information processing, we have shown that NfC *does* moderate the reliance upon policy information. Our results indicate that the null relationship found in past work is most due to regression dilution. We show that moving from a 2-item measure to a 4-item measure items leads to findings that are generally in line with the theoretical expectations derived from the Elaboration Likelihood Model.

Turning to the Big Five and ideology literature, we have shown that – with the exception of Openness and Neuroticism – the association between personality and political dimensions is highly conditional upon the measurement of personality. We found that the BFI tends to underestimate the association between personality and political ideology, while the more elaborate 50-item IPIP-FFM tend to result in stronger associations with political ideology. For the most part, brief measures of personality tend to underestimate the association between personality and political ideology, generally by a factor of two. 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. Moreover, our study confirms that the move from from two-items per trait towards personality measures using a couple more items in the study of personality and political ideology already is an improvement. This is especially the case among traits that have been largely disregarded as relevant for political ideology, namely: Agreeableness, Extraversion, and Conscientiousness. The current study indicates that we should critically assess the current state-of-the-art and reconsider whether brief measures of personality have yielded the right insights into the association between personality and political ideology.

Our study also has three important methodological implications. First, we have

shown that that short measures may yield both Type M and Type S errors by both over-estimating and underestimating the associations between personality and politics. [Gosling et al. \(2003\)](#) already raised awareness that the brief measures such as the TIPI and BFI come at the cost of the psychometric properties. With the currently increasing interest in the association between personality and political ideology, our study raises awareness that the reliance upon brief measures of personality in omnibus surveys has potential detrimental effects on the development of the literature on personality and politics. Scholars relying upon the BFI and the brief NfC at best commit a type M error whereby they underestimate the association with measures of political ideology. A consequence of a type M error is that they conclude that a trait is unrelated to politics. This is clearly seen in the literature with the development of the NfC within political science as well as the importance of some of the Big Five traits that have disregarded for politics.

Second, [Cronbach and Meehl \(1955, 300\)](#) pointed out that: “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.” In the current personality and politics scholars follow [Cronbach and Meehl \(1955\)](#) and assess the internal consistency (see for instance, [Gerber et al., 2010](#); [Mondak, 2010](#)) and over-time stability of the construct ([Gerber et al., 2013](#)). However, limited attention is given to the criterion validity of the construct – but see for notable exceptions recent work on risk taking ([Kam, 2012](#)), imaginativeness ([Petersen & Aarøe, 2013](#)) and disgust sensitivity ([Kam & Estes, 2016](#)). Instead scholars disregard a construct as relevant to politics when their brief measure is not associated with a criterion measure. This paper indicates that ignoring criterion validity may yield incorrect conclusions.

Third, given the current academic concerns over the robustness of findings in social science research, emphasis is given to accessibility of data ([Lupia & Elman, 2014](#)), the size and quality of the sample ([Berinsky et al., 2012](#); [Mullinix et al., 2015](#)), null hypothesis testing ([Gill, 1999](#)) and the preregistration of studies ([Mullinix et al., 2015](#)).

Yet, the impact of inadequate measurement on the criterion validity and thereby enhancing Type M and Type S is not as central to the recent debate. This should be particularly concerning in the domain of personality and politics, where researchers are – with some exceptions (i.e., Bakker, Rooduijn, & Schumacher, 2016; Gallego & Pardos-Prado, 2014; Gerber, Huber, Doherty, & Dowling, 2011) – use brief measures to multidimensional constructs. All of these increase the risk of both Type M and Type S errors (Credé et al., 2012) and influences the conclusions we draw about the role of personality in politics.

This study offers opportunities for further research. First, 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 (see for an example in the personality and politics literature, 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. However, and following Cronbach and Meehl (1955) 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 .

Second, we raise awareness that when it comes to the Big Five, different measurement traditions exist within personality psychology. The IPIP can be considered as part of a more question based approach, while the BFI fits within the adjective approach (John, Naumann, & Soto, 2008). One could argue that differences in criterion validity between the BFI and the IPIP are actually not explained by the length of the battery but by the measurement tradition. Our study suggests that the measurement tradition does not condition the results as our analyses of two item Big Five items based upon the IPIP also suggest that poor measurement . Yet, in order to

assess this alternative explanation in full detail future research one should collect data that contains measures of the BFI, a longer adjective – for instance the 44-item BFI – and the 50-item IPIP.

Third, our study focused upon the NfC and the BFI, which are among the most common brief measures of personality. Yet, future research should also study whether other brief personality – such as the Five Item Personality Inventory employed in the German Election Studies (Schoen & Steinbrecher, 2013), the Big Five Inventory Short Version included in the British and German Household Panels (Hahn, Gottschling, & Spinath, 2012) or the 3-item version of the NfC included in the recent German Election Studies. This is important because there is no a priori reason to assume that these measures are not as prone to the Type M - and to some extent Type S – errors documented in this study, while these brief measures are also employed in the study of personality and politics (Ackermann & Ackermann, 2015; Ackermann et al., 2016; Bakker, Hopmann, & Persson, 2015; Bakker, Klemmensen, Nørgaard, & Schumacher, 2016; ?).

Fourth, we conducted our study in only one political context, namely the Netherlands. The Netherlands is a western European country increasingly used as a political context in the field of personality and politics (see for instance, Bakker & de Vreese, 2016; Bakker, Rooduijn, & Schumacher, 2016; Gallego & Pardos-Prado, 2014; Malka, Soto, Inzlicht, & Lelkes, 2014). While we do not have a priori reasons to expect that the effects should differ in other western countries, it certainly is possible that certain short measures are more valid in one context over another.

Fifth, our research also supports the reconsideration. For instance, scholars could follow Allison and Hauser (1991) and randomly vary the subsets of items that are asked of each respondent and analyzing covariance structures with random missingness. An second alternative is to turn to adaptive tests of personality (Montgomery & Cutler, 2013; Stark, Chernyshenko, Drasgow, & White, 2012). This more dynamic way of measuring constructs saves considerable space in surveys. The reason these scholars propose adaptive tests is that brief measures not capture the latent construct. In this

study, we have provided evidence for the necessity of having *better* measures of personality in large omnibus surveys.

What is the way forward? We have two recommendations for the future of the personality and politics literature in political science. Based upon our study, we strongly advise against employing the 2-item ANES measure to study the reliance upon policy information. Regression dilution seems to seriously affect our results and the conclusions we draw. Instead, scholars are well advised to rely upon larger measures of the NfC such as the 6-item measure employed by [Bullock \(2011\)](#).

The recommendations for the Big Five traits differ. The use of personality traits within omnibus surveys - such as the WVS (for a recent example, see [Fatke, 2016](#)) - should be interpreted with extreme caution. 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.

Besides, we signal that the current state-of-the-art has followed the agenda-setting paper by [Mondak and Halperin \(2008, p. 339\)](#) and continued with “a sweeping search” for the associations between all Big Five traits and political outcomes. In order to do so, scholars decide to measure all traits and thereby compromise on the measure of each of the five traits. We have shown that, compared to brief Big Five measures, slightly more elaborate measures – such as the 20-item Mini-IPIP – yield already fewer Type M errors compared to the 10-item BFI. We understand that time and space are costly – especially in omnibus surveys. However, we believe the literature has arrived at a turning point and should move beyond the use of extremely abbreviated scales of personality.<sup>12</sup>

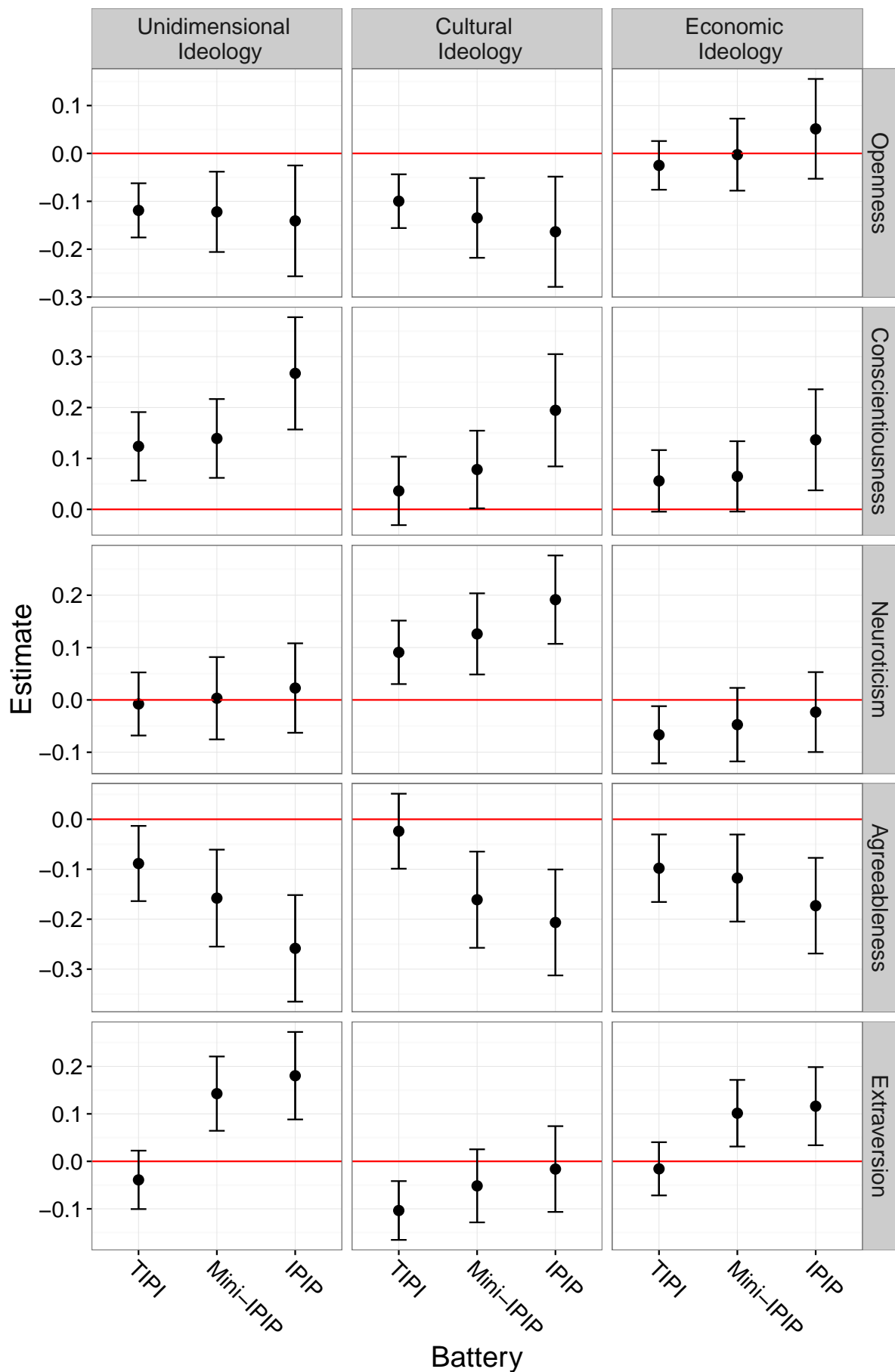
To conclude, we call upon researchers, managers of omnibus surveys and funding

---

<sup>12</sup>Researchers might consider three steps: (1) form theoretically informed expectations about some personality traits and outcomes of interest; (2) measure the traits of interest with more elaborate measures (3) include briefer measures of the other traits in order to account for the fact that the Big Five traits are not orthogonal. Yet, when it comes to *cheaper* forms of data-collection such as MTurk or student samples we see no hurdles for simply implementing more elaborate personality measures such as the 50-item IPIP-FFM. Accordingly, we urge scholars to refrain from using very brief measures of personality in these *cheap* and *easy* accessible samples.

agencies to stop collectively selling ourselves short when using abbreviated measures of the NfC and the Big Five. For the development of personality and politics literature it is time to include measure of only a few of individual differences, instead of aiming to measure as much constructs as possible. The trade-offs between different constructs might be a tough one but we have documented that the consequences of using brief measures out-weight the gains of spacing time and space in the survey. Accordingly, our study hopefully sets the stage for the next generation of research in political science.

Figure 3. Personality and Politics: BFI, Mini-IPIP and IPIP-FFM results

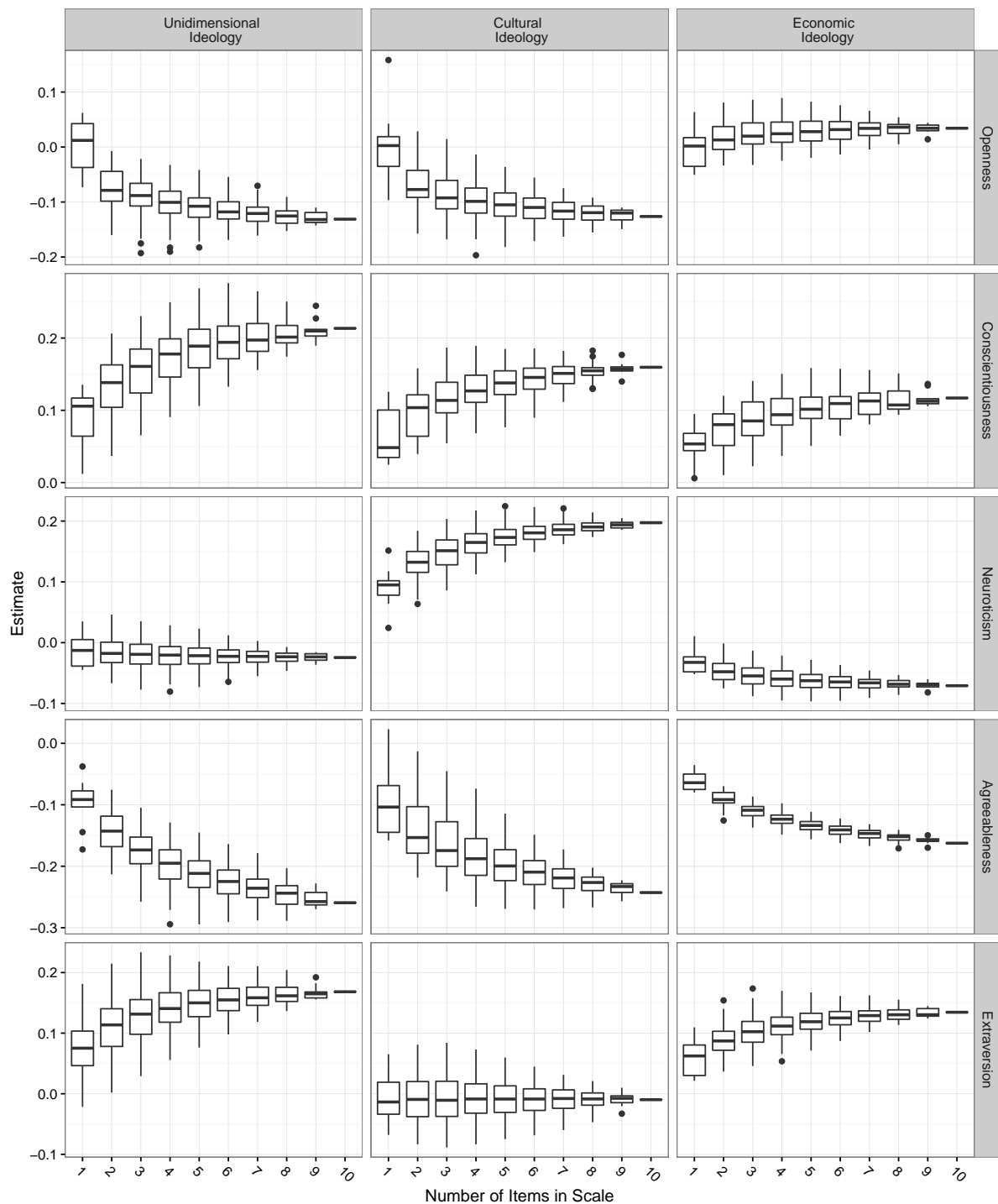


Note: OLS estimates with 95% Confidence Intervals are plotted. The Supplementary

Material provides the tables with results of the regression models for the models



Figure 4. Personality and Political Ideology:s



Figures plots the associations between our measures of ideology and each of these 1022 possible combinations of each trait. Distribution of the point estimates of these measures sorted by the number of items used to generate the trait. We plot distribution of point estimates of the regression coefficients from models predicting each form of ideology (the columns) to each 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

## References

- Ackermann, K., & Ackermann, M. (2015). The big five in context: Personality, diversity and attitudes toward equal opportunities for immigrants in Switzerland. *Swiss Political Science Review*, 21(3), 396–418.
- Ackermann, K., Ackermann, M., & Freitag, M. (2016). Opting for an open society? Personality traits and attitudes toward the openness of Switzerland. *Comparative European Politics*.
- Allison, P. D., & Hauser, R. M. (1991). Reducing bias in estimates of linear models by remeasurement of a random subsample. *Sociological Methods & Research*, 19(4), 466–492.
- 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*, 1055–1082.
- Bakker, B. N. (2016). Personality traits, income and economic ideology. *Political Psychology*. doi: 10.1111/pops.12349
- Bakker, B. N., & de Vreese, C. H. (2016). Personality and EU attitudes: Relationship across EU attitudes. *European Union Politics*, 1(17), 25–45.
- Bakker, B. N., Hopmann, D. N., & Persson, M. (2015). Personality traits and party identification over time. *European Journal of Political Research*, 54, 197–215.
- Bakker, B. N., Klemmensen, R., Nørgaard, A. S., & Schumacher, G. (2016). Stay loyal or exit the party? how openness to experience and extroversion explain vote switching. *Political Psychology*, 37(3), 419–429.
- 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., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351–368.
- 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(03), 496–515.
- Burisch, M. (1984a). Approaches to personality inventory construction: A comparison of merits. *American Psychologist*, 39(3), 214.
- Burisch, M. (1984b). You don't always get what you pay for: Measuring depression with short and simple versus long and sophisticated scales. *Journal of Research in Personality*, 18(1), 81–98.
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116.
- 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.
- Carney, D. R., Jost, J. T., Gosling, S. D., & Potter, J. (2008). The secret lives of liberals and conservatives: Personality profiles, interaction styles, and the things they leave behind. *Political Psychology*, 29(6), 807–840.
- Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from mechanical turk valid for research on political ideology? *Research & Politics*, 2(4), 2053168015622072.
- Coffe, H. (2013). It's all about the party: Gender, party characteristics, and radical right voting.
- 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), 874–88.

- Cronbach, L. J. (1949). Essentials of psychological testing.
- Cronbach, L. J., & Gleser, G. C. (1957). Psychological tests and personnel decisions.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281.
- De Neve, J.-E. (2015). Personality, childhood experience, and political ideology. *Political Psychology*, 36(1), 55–73.
- Dietrich, B. J., Lasley, S., Mondak, J. J., Rempel, M. L., & Turner, J. (2012). Personality and legislative politics: The big five trait dimensions among us state legislators. *Political Psychology*, 33(2), 195–210.
- Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-ipp scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18(2), 192.
- 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.
- Ehrhart, M. G., Ehrhart, K. H., Roesch, S. C., Chung-Herrera, B. G., Nadler, K., & Bradshaw, K. (2009). Testing the latent factor structure and construct validity of the ten-item personality inventory. *Personality and Individual Differences*, 47(8), 900–905.
- Fatke, M. (2016). Personality traits and political ideology: A first global assessment. *Political Psychology*.
- Feldman, S., & Johnston, C. (2014). Understanding the determinants of political ideology: Implications of structural complexity. *Political Psychology*, 35(3), 337–358.
- Freitag, M., & Rapp, C. (2015a). The personal foundations of political tolerance towards immigrants. *Journal of Ethnic and Migration Studies*, 41(3), 351–373.
- Freitag, M., & Rapp, C. (2015b). The personal foundations of political tolerance towards immigrants. *Journal of Ethnic and Migration Studies*, 41(3), 351–373.
- Friesen, A., & Hibbing, M. V. (2016). The effect of personal economic values on

- economic policy preferences. *Social Science Quarterly*.
- Furnham, A. (2008). Relationship among four big five measures of different length 1. *Psychological Reports, 102*(1), 312–316.
- Gallego, A., & Pardos-Prado, S. (2014). The big five personality traits and attitudes towards immigrants. *Journal of Ethnic and Migration Studies, 40*(1), 79–99.
- 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*(01), 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*(03), 692–706.
- Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political Research Quarterly, 52*(3), 647–674.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe, 7*(1), 7–28.
- 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.
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's 'ipip' big-five factor markers: Internal consistency and concurrent validation in scotland. *Personality and Individual Differences*, 39(2), 317–329.
- Green, D. P. (1988). On the dimensionality of public sentiment toward partisan and ideological groups. *American Journal of Political Science*, 758–780.
- Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, 64(6), 1029.
- Ha, S. E., Kim, S., & Jo, S. H. (2013). Personality traits and political participation: Evidence from south korea. *Political Psychology*, 34(4), 511–532.
- Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of personality—validity and reliability of the gsoep big five inventory (bfi-s). *Journal of Research in Personality*, 46(3), 355–359.
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). Differences in negativity bias underlie variations in political ideology. *Behavioral and Brain Sciences*, 37(03), 297–307.
- Hofmans, J., Kuppens, P., & Allik, J. (2008). Is short in length short in content? an examination of the domain representation of the ten item personality inventory scales in dutch language. *Personality and Individual Differences*, 45(8), 750–755.
- Holbrook, T. M. (2006). Cognitive style and political learning in the 2000 us presidential campaign. *Political Research Quarterly*, 59(3), 343–352.
- John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology*, 60(3), 348.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research*, 3, 114–158.
- Johnson, T. R., & Martin, A. D. (1998). The public's conditional response to supreme court decisions. *American Political Science Review*, 92(02), 299–309.
- Johnston, C. D. (2013). Dispositional sources of economic protectionism. *Public*

- Opinion Quarterly*, 77(2), 574–585.
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339–375.
- Jost, J. T., West, T. V., & Gosling, S. D. (2009). Personality and ideology as determinants of candidate preferences and “obama conversion” in the 2008 us presidential election. *Du Bois Review: Social Science Research on Race*, 6(01), 103–124.
- Kam, C. D. (2005). Who toes the party line? cues, values, and individual differences. *Political Behavior*, 27(2), 163–182.
- Kam, C. D. (2012). Risk attitudes and political participation. *American Journal of Political Science*, 56(4), 817–836.
- Kam, C. D., & Estes, B. A. (2016). Disgust sensitivity and public demand for protection. *The Journal of Politics*, 78(2), 481–496.
- Ksiazkiewicz, A., Ludeke, S., & Krueger, R. (2016). The role of cognitive style in the link between genes and political ideology. *Political Psychology*.
- Leeper, T. J. (2014). Cognitive style and the survey response. *Public Opinion Quarterly*, 78(4), 974–983.
- Leeson, P., & Heaven, P. C. L. (1999). Social attitudes and personality. *Australian Journal of Psychology*, 5(I), 19–24.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley Publishing Company.
- Lupia, A., & Elman, C. (2014). Openness in political science: Data access and research transparency. *PS: Political Science & Politics*, 47(01), 19–42.
- Lyons, J., Sokhey, A. E., McClurg, S. D., & Seib, D. (2016). Personality, interpersonal disagreement, and electoral information. *The Journal of Politics*, 78(3), 000–000.
- Malka, A., Soto, C. J., Inzlicht, M., & Leles, 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.

- McClosky, H. (1958). Conservatism and personality. *American Political Science Review*, 52(01), 27–45.
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23(1), 1–21.
- Mérola, V., & Hitt, M. P. (2015). Numeracy and the persuasive effect of policy information and party cues. *Public Opinion Quarterly*, nfv051.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741.
- Mondak, J. J. (2010). *Personality and the foundations of political behavior*. Cambridge University Press.
- 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(01), 85–110.
- Montgomery, J. M., & Cutler, J. (2013). Computerized adaptive testing for public opinion surveys. *Political Analysis*, 21(2), 172–192.
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(02), 109–138.
- Nielsen, J. H. (2016). Personality and euroscepticism: The impact of personality on attitudes towards the eu. *JCMS: Journal of Common Market Studies*.
- Niemi, R. G., Carmines, E. G., & McIver, J. P. (1986). The impact of scale length on reliability and validity. *Quality and Quantity*, 20(4), 371–376.
- Osborne, D., & Sibley, C. G. (2012). Does personality matter? openness correlates with vote choice, but particularly for politically sophisticated voters. *Journal of Research in Personality*, 46(6), 743–751.
- Osborne, D., & Sibley, C. G. (2015). Within the limits of civic training: Education



- moderates the relationship between openness and political attitudes. *Political Psychology*, 36(3), 295–313.
- Paunonen, S. V., & Jackson, D. N. (1985a). The validity of formal and informal personality assessments. *Journal of Research in Personality*, 19(4), 331–342.
- Paunonen, S. V., & Jackson, D. N. (1985b). The validity of formal and informal personality assessments. *Journal of Research in Personality*, 19(4), 331–342.
- Petersen, M. B., & Aarøe, L. (2013). Politics in the mind's eye: Imagination as a link between social and political cognition. *American Political Science Review*, 107(02), 275–293.
- 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.
- Rentfrow, P. J., Gosling, S. D., Jokela, M., Stillwell, D. J., Kosinski, M., & Potter, J. (2013). Divided we stand: Three psychological regions of the united states and their political, economic, social, and health correlates. *Journal of personality and social psychology*, 105(6), 996.
- Rentfrow, P. J., Jost, J. T., Gosling, S. D., & Potter, J. (2009). Statewide differences in personality predict voting patterns in 1996–2004 us presidential elections. *Social and Psychological Bases of Ideology and System Justification*, 1, 314–349.
- Riemann, R., Grubich, C., Hempel, S., Mergl, S., & Richter, M. (1993). Personality and attitudes towards current political topics. *Personality and Individual Differences*, 15(3), 313–321.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the rosenberg self-esteem scale. *Personality and Social Psychology Bulletin*, 27(2), 151–161.
- Rojas, S. L., & Widiger, T. A. (2014). Convergent and discriminant validity of the five factor form. *Assessment*, 21(2), 143–157.
- Rudolph, T. J. (2011). The dynamics of ambivalence. *American Journal of Political Science*, 55(3), 561–573.

- Sadowski, C. J., & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the need for cognition scale. *Perceptual and Motor Skills, 74*(2), 610–610.
- 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*.
- Schoen, H., & Steinbrecher, M. (2013). Beyond total effects: Exploring the interplay of personality and attitudes in affecting turnout in the 2009 german federal election. *Political Psychology, 34*(4), 533–552.
- Sibley, C. G., Osborne, D., & Duckitt, J. (2012). Personality and political orientation: Meta-analysis and test of a Threat-Constraint model. *Journal of Research in Personality, 46*(6), 664–677.
- Silvia, P. J., & Sanders, C. E. (2010). Why are smart people curious? fluid intelligence, openness to experience, and interest. *Learning and Individual Differences, 20*(3), 242–245.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*(1), 102.
- Smits, N., & Vorst, H. C. (2007). Reducing the length of questionnaires through structurally incomplete designs: An illustration. *Learning and Individual Differences, 17*(1), 25–34.
- Sniderman, P. M. (1975). *Personality and democratic politics*. Univ of California Press.
- Sokhey, A. E., & McClurg, S. D. (2012). Social networks and correct voting. *The Journal of Politics, 74*(03), 751–764.
- Stark, S., Chernyshenko, O. S., Drasgow, F., & White, L. A. (2012). Adaptive testing with multidimensional pairwise preference items: Improving the efficiency of personality and other noncognitive assessments. *Organizational Research Methods, 1094428112444611*.
- Sweetser, K. D. (2013, October). Partisan Personality: The Psychological Differences Between Democrats and Republicans, and Independents Somewhere in Between. *American Behavioral Scientist*. doi: 10.1177/0002764213506215

- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011a). Comparative validity of brief to medium-length big five and big six personality questionnaires. *Psychological assessment, 23*(4), 995.
- Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011b). Comparative validity of brief to medium-length big five and big six personality questionnaires. *Psychological Assessment, 23*(4), 995.
- Van Hiel, A., & Mervielde, I. (2004). Openness to experience and boundaries in the mind: Relationships with cultural and economic conservative beliefs. *Journal of Personality, 72*(4), 659–686.
- Verhulst, B., Eaves, L. J., & Hatemi, P. K. (2012). Correlation not causation: The relationship between personality traits and political ideologies. *American Journal of Political Science, 56*(1), 34–51.
- Verhulst, B., Hatemi, P. K., & Martin, N. G. (2010). The nature of the relationship between personality traits and political attitudes. *Personality and Individual Differences, 49*(4), 306–316.
- Weinschenk, A. C. (2014). Personality traits and the sense of civic duty. *American Politics Research, 42*(1), 90–113.
- Woods, S. A., & Hampson, S. E. (2005). Measuring the big five with single items using a bipolar response scale. *European Journal of Personality, 19*(5), 373–390.
- Ziegler, M., Kemper, C. J., & Kruey, P. (2014). Short scales—five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*(4), 185–189.