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

\footnote{We thank Martijn Schoonvelde 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

Brief measures of personality are increasingly included in large omnibus surveys where space and time are limited. The trade-off for saving space might be the validity of our findings. We investigate the impact of short measures on the conclusions we draw about the link between personality and politics by examining two central questions in the literature. The first study investigates whether Need for Cognition (NfC) moderates the impact of political information on attitudes. When we use the popular 2-item measure of NfC, we find that NfC does not impact cue taking, which is in line with most extant political science research. When we use larger measures, we find a substantive and significant effect, which is in line with theories of information processing. In the second study, we investigate whether the Big Five personality traits are related to ideology. We show that the more elaborate measures of personality often yield relationships that are about twice as large as those found when we use the ubiquitous Ten Item Personality Inventory (TIPI). Traits that been largely dismissed as relevant to political ideology – Neuroticism, Agreeableness, Extraversion and even Conscientiousness – yield stronger and more consistent associations when larger batteries are employed versus the brief TIPI. Both studies imply that the measurement of personality conditions the conclusions we draw about the role of personality in politics. (Abstract is 226 words).

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), Leeper (2014), 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 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 TIPI, 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. Relatively, 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 TIPI, the NfC as well as other brief personality measures are satisfactory. The test-retest reliability of the 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 the TIPI was compared to the personality batteries with a modest length the convergent validity was acceptable (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 (Kruyen, Emons, & Sijtsma, 2013). First, in all domains of research, it’s well known that short batteries tend to be less reliable than longer batteries (Lord & Novick, 1968),
and 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).²

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 1960, 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 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 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 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

²Niemi 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.
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.\(^3\) 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.\(^4\) 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.

**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).

\(^3\)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, 2011).

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

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) re-scaled 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 male or female right-wing political candidate. Respondents either saw 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.” 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 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 percent response rate) and May 2011

---

5See Appendix X for full text of all experimental treatments.
SELLING OURSELVES SHORT? 10

(74 percent 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 re-scaled 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 indicates that three NfC batteries are strongly correlated.

Table 1

<table>
<thead>
<tr>
<th>Battery</th>
<th>Mean</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>α</th>
<th>r Bullock</th>
<th>r ANES</th>
</tr>
</thead>
<tbody>
<tr>
<td>full 18-item</td>
<td>0.55</td>
<td>0.15</td>
<td>-0.13</td>
<td>3.05</td>
<td>.90</td>
<td>.91</td>
<td>.78</td>
</tr>
<tr>
<td>Bullock 6-item</td>
<td>0.55</td>
<td>0.18</td>
<td>-0.09</td>
<td>2.81</td>
<td>.77</td>
<td>1</td>
<td>.83</td>
</tr>
<tr>
<td>ANES 2-item</td>
<td>0.54</td>
<td>0.24</td>
<td>-0.32</td>
<td>2.55</td>
<td>.84</td>
<td>.83</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note:* N = 4,733

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, 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. Specifically, 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 re-scaled 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 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.
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 C).

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.

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, 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, & Kruyen, 2014), another measure of reliability that is less dependent upon the length of the scale (McDonald, 1970). Yet, we show in the Supplementary Materials that a brief scale

\footnote{Omega can only be calculated when the scale contains three items or more.}
SELLING OURSELVES SHORT?

with a higher Omega coefficient does not lead to more reliable estimates compared to a brief scale with a lower Omega coefficient. 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?

**Short measures with high bandwidth: The Case of the TIPI**

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. The Ten Item Personality Inventory, which measures each trait using two items (TIPI, Gosling et al., 2003), has been perhaps the most popular personality battery used in recent political science research. The TIPI is 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.7

The TIPI measures the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism), which are each high bandwidth constructs. The items for a short measure of each trait are selected so that the short

---

7Yougov - which is one the most prominent pollsters - includes the TIPI in their panel as part of their general background variables.
measure reflects the breadth of the original dimension. Accordingly, the inter-item correlation will be low (Gosling et al., 2003; Woods & Hampson, 2005), necessarily decreasing the reliability and criterion validity of the scale. Second, and more fundamentally, all personality indices 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 – each trait 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, Remmel, & Turner, 2012; Freitag & Rapp, 2015; Friesen & Hibbing, 2016; Gerber, Huber, Doherty, & Dowling, 2011; Gerber et al., 2010; Ha, Kim, & Jo, 2013; Jost, West, & Gosling, 2009; Rentfrow et al., 2013; Rentfrow, Jost, Gosling, & Potter, 2009; Sweetser, 2013,?; Verhulst, Hatemi, & Martin, 2010; 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; Verhulst et al., 2010). Scholars have also consistently found no relationship between Openness and economic conservatism (Bakker, 2016; Carney et al., 2008; 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 et al., 2010), while others, using the TIPI, found positive associations (Gerber, Huber, Doherty, & Dowling, 2011; Gerber et al., 2010). Similarly, there is some evidence that Neuroticism correlates negatively with economic conservatism (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 their 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). 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 TIPI 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 TIPI (Ehrhart et al., 2009). 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), which has recently also been used in the study of personality and politics (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 TIPI 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.8 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.9

---

8This means the survey was conducted in Dutch.
9This means we have a data-set that includes 84% of the respondents that completed the WVS 2012 in the Netherlands.
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., TIPI), 4-items (i.e., Mini-IPIP) and 10-items (i.e., IPIP-FFM) per trait. The TIPI was administered as part of the WVS 2012 in the Netherlands. The Dutch-translation of the TIPI has been shown to be a valid and reliable measure of personality (Hofmans et al., 2008).

The TIPI 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 TIPI compared to the measures collected earlier.

For each of the personality inventories employed in this study, we created additive indices 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 – TIPI 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.\textsuperscript{10} 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 TIPI traits than the Mini-IPIP and IPIP-FFM, the means were very similar. The correlations between the TIPI 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 constructed based upon the items of the full IPIP.

\textbf{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).

\textit{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).

\textit{Cultural ideology} was also 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).

\textsuperscript{10}The internal consistency of the TIPI is comparable to other studies using population based samples.
Table 2

*Descriptive statistics*

<table>
<thead>
<tr>
<th>Trait</th>
<th>Battery</th>
<th>Mean</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>α</th>
<th>r</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openness</td>
<td>TIPI</td>
<td>0.54</td>
<td>0.25</td>
<td>-0.08</td>
<td>2.34</td>
<td>0.46</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mini-IPIP</td>
<td>0.59</td>
<td>0.18</td>
<td>0.04</td>
<td>3.09</td>
<td>0.60</td>
<td>0.47</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>IPIP-FFM</td>
<td>0.57</td>
<td>0.14</td>
<td>0.07</td>
<td>2.95</td>
<td>0.76</td>
<td>0.45</td>
<td>0.84</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>TIPI</td>
<td>0.73</td>
<td>0.21</td>
<td>-0.56</td>
<td>2.66</td>
<td>0.42</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mini-IPIP</td>
<td>0.67</td>
<td>0.18</td>
<td>-0.39</td>
<td>2.83</td>
<td>0.67</td>
<td>0.46</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>IPIP-FFM</td>
<td>0.68</td>
<td>0.13</td>
<td>-0.29</td>
<td>2.93</td>
<td>0.77</td>
<td>0.51</td>
<td>0.89</td>
</tr>
<tr>
<td>Extraversion</td>
<td>TIPI</td>
<td>0.53</td>
<td>0.23</td>
<td>0.19</td>
<td>2.64</td>
<td>0.41</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mini-IPIP</td>
<td>0.55</td>
<td>0.19</td>
<td>-0.07</td>
<td>2.84</td>
<td>0.75</td>
<td>0.46</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>IPIP-FFM</td>
<td>0.54</td>
<td>0.17</td>
<td>-0.01</td>
<td>2.92</td>
<td>0.87</td>
<td>0.51</td>
<td>0.93</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>TIPI</td>
<td>0.69</td>
<td>0.18</td>
<td>-0.40</td>
<td>3.09</td>
<td>0.19</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mini-IPIP</td>
<td>0.72</td>
<td>0.16</td>
<td>-0.60</td>
<td>3.66</td>
<td>0.76</td>
<td>0.22</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>IPIP-FFM</td>
<td>0.68</td>
<td>0.15</td>
<td>-0.51</td>
<td>3.81</td>
<td>0.83</td>
<td>0.24</td>
<td>0.91</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>TIPI</td>
<td>0.35</td>
<td>0.24</td>
<td>0.42</td>
<td>2.61</td>
<td>0.68</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mini-IPIP</td>
<td>0.34</td>
<td>0.19</td>
<td>0.39</td>
<td>3.01</td>
<td>0.73</td>
<td>0.62</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>IPIP-FFM</td>
<td>0.37</td>
<td>0.17</td>
<td>0.38</td>
<td>2.91</td>
<td>0.88</td>
<td>0.67</td>
<td>0.93</td>
</tr>
</tbody>
</table>

*Note: N = 1582*

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 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*.\(^{11}\) To make the results easily comparable, we plot 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 TIPI, 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 TIPI (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 TIPI 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.”\(^{12}\)

\(^{11}\)The results do not change if we do not control for education and income.

\(^{12}\)Note that recent research suggest that the association between personality and economic ideology might be moderated by political interest (Federico, Johnston, & Lavine, 2014) or income (Bakker, 2016).
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 TIPI 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 TIPI 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 TIPI. Columns 2 and 3 of Figure 3, row 2, indicate that scholars using the TIPI 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 3, 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 TIPI (b=0.10, SE=0.05, p<0.05). The TIPI 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 TIPI (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 TIPI, while the Mini-IPIP estimate is roughly 50% larger compared to the TIPI although not statistical significant different. The TIPI 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 statistical significant ($b=-0.18$, $SE=0.06$, $p<0.05$) – than the TIPI estimate. Finally, if we employ the TIPI in the study of economic ideology, we would likely conclude that the 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 TIPI and thus commit a Type M error.

Extraversion yielded the most striking results with both Type M and Type S errors occurring. Using the TIPI, 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 statistical significantly stronger ($b=0.20$, $SE=0.06$, $p<0.05$) compared to the TIPI estimate. The Mini-IPIP estimate was four times larger than the TIPI estimate. While those using the TIPI 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 TIPI 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 (Gosling et al., 2003), Single Item Measure of Personality (Woods & Hampson, 2005), Big Five Inventory 10-item (Rammstedt & John, 2007) – would results 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 18-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.¹³

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

---

¹³Graham et al. (2011) use a similar approach to assess the consequences of using brief versions of the Moral Foundation Questionnaire.
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 TIPI’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 TIPI.

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 TIPI 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 substantive and methodological contribution to the personality and politics literature. 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 politics 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 TIPI 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 TIPI 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 shows 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 comes 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 TIPI 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 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, in their seminal study, Cronbach and Meehl (1955) 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.\textsuperscript{14} 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 et al., 2016; De Neve, 2015; Gerber, Huber, Doherty, & Dowling, 2011; Hatemi & Verhulst, 2015) – 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.\textsuperscript{15} This implies that we have to be careful in drawing conclusions.

\textsuperscript{14}A notable exception is the work by Kam on risk taking (Kam, 2012) and disgust sensitivity (Kam & Estes, 2016).
\textsuperscript{15}The Netherlands does not have validated measures of turn-out as in the U.S.
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 have mapped the consequences of using brief Big Five and NfC measures and showed that adding a couple of items to a brief measure already seem to improve the estimate remarkably. It was, however, beyond the scope of this study to isolate and validate the exact number of items that would yield the best estimate. Future research - using independent samples - could (1) replicate our findings and (2) isolate the items that make up the most efficient measure of NfC or any Big Five trait.

Third, 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 TIPI fits within the adjective approach (John, Naumann, & Soto, 2008). One could argue that differences in criterion validity between the TIPI 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 TIPI, a longer adjective – for instance the 44-item BFI – and the 50-item IPIP.

Fourth, our study focused upon the NfC and the TIPI, 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 (Donnellan & Lucas,
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, 2015; ?). Accordingly, this study only marks the starting point of the assessment of the consequences of the use of brief measures of personality.

Fifth, 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 et al., 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.

What is the way forward? We have three recommendations for the future of the personality and politics literature in political science. Based upon our study, we 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 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 TIPI. However, we understand that time and space are costly – especially in omnibus surveys.
Therefore, researchers might consider these 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 brief measures of the other traits in order to account for the fact that the Big Five traits are not orthogonal.\(^{16}\)

To conclude, we call upon researchers, managers of omnibus surveys and funding agencies to more critically assess the extent to which we are not collectively selling ourselves short when using abbreviated measures of the NfC and the Big Five. For the development of the theoretical understanding of the role of personality in politics, it seems better 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 thought 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 paper hopefully will set the stage for the next generation of research in political science.

\(^{16}\)When it comes to \textit{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 \textit{cheap} and \textit{easy} accessible samples.
Figure 3. Personality and Politics: TIPI, 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 including the TIPI (Table 3), the Mini-IPIP (Table 4) and the IPIP-FFM (Table 5).
Figure 4. Personality and Political Ideology:

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


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


*Political Psychology, 36*(1), 55–73.


*Psychological assessment, 18*(2), 192.


Psychological Reports, 102(1), 312–316.


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.


big-five factor markers: Internal consistency and concurrent validation in scotland. 


Ha, S. E., Kim, S., & Jo, S. H. (2013). Personality traits and political participation: 

Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of 

Hatemi, P. K., & Verhulst, B. (2015). Political attitudes develop independently of 
personality traits. *PloS one, 10*(3), e0118106.

underlie variations in political ideology. *Behavioral and Brain Sciences, 37*(03), 
297–307.

examination of the domain representation of the ten item personality inventory 


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 

five trait taxonomy. *Handbook of Personality: Theory and Research, 3*, 114–158.


Opinion Quarterly, 77*(2), 574–585.


Paunonen, S. V., & Jackson, D. N. (1985b). The validity of formal and informal


SELLING OURSELVES SHORT? 41


