Supporting Information/Online Appendix for *Legislators Off Their Leash: Cognitive Shirking and Impending Retirement in the U.S. Congress*

Data Collection

As discussed in the main text, the data here is a collection of all representatives whom sought retirement between 1997 and 2013 (the 105th to the 112th Congresses, respectively). As the analysis here is specifically interested in how representatives may cognitively change their behavior once they have decided to exit electoral politics, the first step in data collection was to create a list of only those members whom decided to exit voluntarily from congress, and to remove any member who chose to leave for alternative reasons. In order to generate a list of retiring representatives, I utilized the "Casualty List," created by Congressional Quarterly at the end of each congressional term.[1] The Roll Call Casualty List was utilized because of its choice to separate those outgoing members based on the rationale behind their decision to leave office, ensuring that only members who specifically chose retirement (over resignation for higher/alternative office, or losing in a primary/general election) were included in the initial list. This list was then verified using a secondary database of outgoing members via the Open Secrets webpage.

Once a list of representatives had been collected, textual data for each member was collected via Congress's THOMAS database. As noted in the text, I focus explicitly on representative's statements made on the floor of the chamber in my examination of cognitive complexity. Obtaining floor speeches for each member followed a

[1] The list can be found publically here: http://data.rollcall.com/media/casualtylists/

straightforward process[2] utilizing a unique python script process.[3] Each member's remarks were carefully coded by the date they were given in the congressional record, and then downloaded into plain-text documents for processing.

LIWC Classifications & Cognitive Complexity

In the main text, the analysis utilizes a host of measures calculated through the LIWC database in order to estimate cognitive complexity. Here, I will give a full explanation for each measure. In total, nine measures of differentiation and integration were utilized in order to form the measure of cognitive complexity used as the primary independent variable in the analysis. One final indicator, the number of words with six or more letters in a speech was also included in the calculation of cognitive complexity. The *six-letter* dimension is commonly used in linguistics to determine a speaker's level of sophistication. Previous examinations of integration and cognitive complexity have shown that increased amounts of six-letter words in a speech are associated with higher levels of cognitive complexity (Gruenfeld 1995; Owens and Wedeking 2012).

Generally, empirical examination of cognition have measured cognitive complexity by first determining the levels of differentiation and integration used a speakers language, and then collapsing these into a single score to measure the extent of a

---

[2] Unfortunately, as of this writing, congress.gov has instituted the use of Cloudflare Security protocols in order to protect the site from Distributed Denial of Service (DDoS) attacks. While this system does provide substantial support for congress.gov, it also makes it more difficult to collect bulk data from a site, making continued research more difficult.

[3] The author upon request can provide this code.

speaker's evaluative and decision-making power. Differentiation focuses on the capacity of an individual to relate to and adopt a variety of perspectives when introduced to a particular event or issue. In their case study of press conferences held by Rudy Guiliani between 1993 and 2001, Pennebaker and Lay (2002) show, for example, that Guiliani began to convey a greater sense of differentiation when considering multiple perspectives and compromise as a result of his personal battle with cancer and the terrorist attacks on September 11. Put plainly then, differentiation examines whether individuals see the world in "black and white" or "shades of gray." The second dimension, integration, takes a step past differentiation and indicates that an individual has the capacity to recognize connections between divergent perspectives. Complex thinkers tend to consider an extensive range of issues and attempt to reconcile contradictions, whereas lower complexity displays rigid adherence to attitudes (Winters 1996).

In the analysis, six of dimensions identified were considered to positively affect cognitive complexity. Complexity, as a unit of measure here, is based upon the ability of a speaker to clearly differentiate between ideas as well as integrate these ideas into cohesive arguments. The clearest example of how this may occur in speech is through the use of terms associated with *causation.* LIWC measures causation by searching a document for terms such as, "because," "effect," "hence," and other terms that refer to causal processes. Causation focuses on the ability of a speaker to tap into and understand the relationships between various parts or components (differentiation), and particularly how changes in one circumstance may influence changes in another (integration). Increased levels of causation typically correspond with greater cognitive complexity overall. Complementary to causation, *inclusiveness* captures the degree to which a speaker sees connections between ideas and concepts, and is measured by LIWC based on the occurrences of terms such as, “with,” and “and.”

Along with causation, *insight* and *discrepancy* work in tandem to indicate the ability of a speaker to provide depth and greater understanding to complex phenomenon. *Insight*, as measured by LIWC, is based on the use of terms such as, “think,” “know,” and “consider” in speech. Insight captures the degree and depth of an individual’s understanding about a experience, as well as capturing how well a speaker understands the underlying nature of an event. Along with this, *discrepancy* measures the use of terms like, “should,” “would,” and “could” and provides details about the ability of a speaker to identify inconsistencies between different cases. Greater levels of discrepancy are indicative that a speaker is able to differentiate between occurrences or events, while the complimentary insight indicates higher levels of integration between complex events.

The final two indicators included in the measure that are positively associated with cognitive complexity calculate the amount of *inhibition* and *tentativeness* a speaker has in their text. Inhibition and tentativeness both examine similar underlying phenomenon that may affect a speaker’s text. A higher degree of inhibition and tentativeness are theorized to indicate how much restraint an individual expresses as well as how they may be impeded in acting. That is, how hesitant is a speaker, or how may they be hindered in acting. *Inhibition* is measured by LIWC by counting occurrences of terms such as, “block,” “stop,” and “constrain.” *Tentativeness* as a dimension includes terms such as, “maybe,” “fairly, “and, “perhaps.”

Along with the aforementioned indices that positively affect cognitive complexity, three indicators were included that are theorized to negatively affect the complexity of a speaker. Specifically, *certainty* is included in the measure by counting the incidents of terms such as, "always, "clearly," or, "absolutely." Certainty is generally associated with the degree of confidence a speaker has in their determination of something, and also tends to be correlated with expressions that paint issues as less complex. As a result, increased certainty tends to diminish a speakers expressed ability to provide integrative understanding or show clear differentiation between courses of action. Along with this, the use of *exclusiveness* in speech is generally used to provide distinction between ideas, and is measured by searching for the use of terms like, "but," or, "except." Speakers tend to use such terms directly when they are determining whether something belongs in one category or not. While the language of exclusiveness can show some clear differentiation, it scores low on showing the ability of a speaker to provide integrative complexity, which is also key to a well-rounded complexity score. LIWC measures the final quantity, *negation*, by searching for terms such as, "no," and, "never." The use of these terms generally is associated to occurrences when a speaker acknowledges the absence of something that is considered positive or affirmative.

In order to determine whether these ten categories represent one underlying concept, the indices were subjected to an exploratory factor analysis using principle components. Based on an analysis of the variance using both the Guttman-Kaiser and Broken Stick model, the results of the factor analysis returned a one-factor solution. Visual representations of these two models are provided in Figure 1. According to the Guttman-Kaiser model, factors should be retained based upon the magnitude of their eigenvalues, with the general view being that values greater than the average should be retained. The bar chart of the eigenvalues at the top of Figure 1 indicate that a three-factor solution is possible for the analysis, however factors two and three only minimally meet the threshold for retaining. As a second test, therefore, the Broken Stick model was implemented, and clearly indicated that a single-factor solution was sufficient. The Broken Stick model is based on an examination of the total percentage of the variance explained by each component. Since the eigenvalue of a PCA represents a measure of each component's variance, components are retained only if the associated value is larger than the value given by the broken stick distribution (MacArthur 1957; Cangelosi and Goriely 2007). These results provide confidence that the ten indicators included are part of the same underlying dimension. Finally, the alpha scale reliability measure for the ten items is .78.

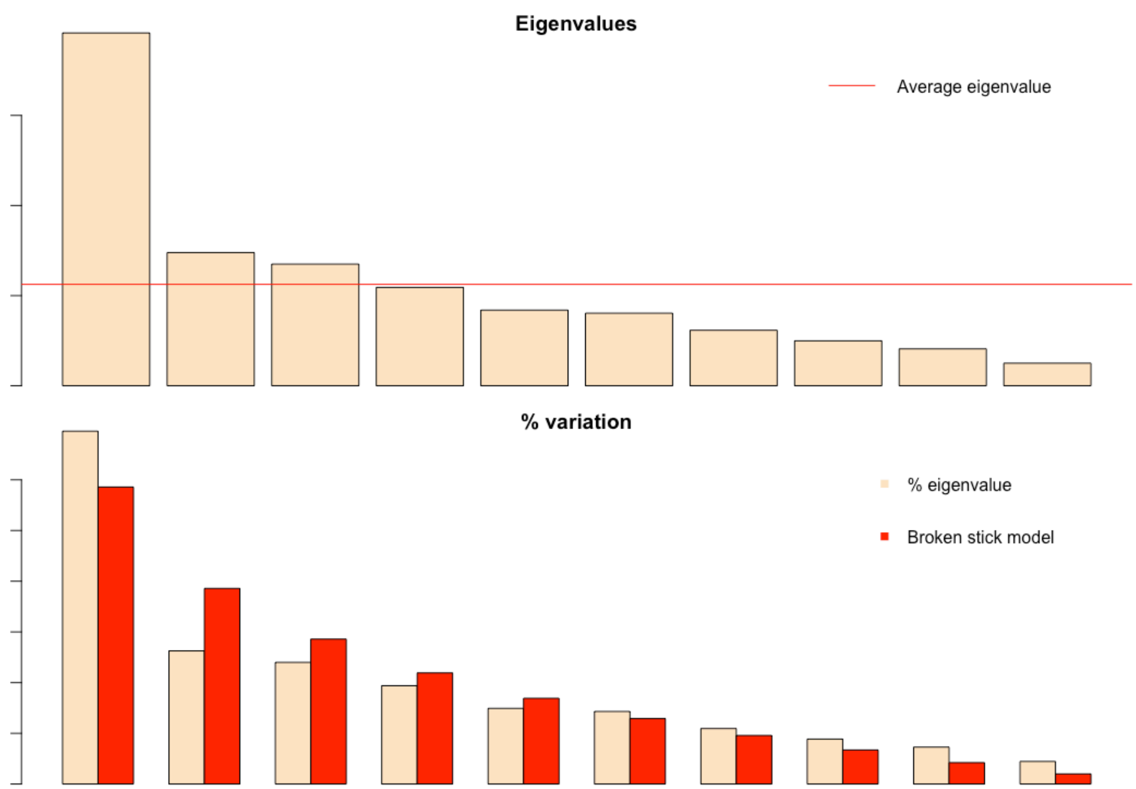

Figure 1: Guttman-Kaiser plot and Broken Stick Plot of the eigenvalues, based on a principle component analysis of the ten factors included in the creation of the cognitive complexity variable used in the main analysis.

In order to create the final variable for cognitive complexity, the indices included were first standardized and a Z-score was produced. To create the final score, I follow a formula first developed by Gruenfeld (1995) and replicated by Owens and Wedeking (2012) in their examinations of cognitive complexity and Supreme Court opinions. Accordingly, Cognitive complexity = *Zsixletter – Zcausation – Zinclusive – Zinsight – Zdiscrepency – Zinhibit – Ztentative – Zcertainty – Zexclusive – Znegation*. To generate the measure of cognitive inconsistency, I simply take the standard deviation for each member of Congress's set of documents.