Supplementary Materials

Additional AutoIC-Human Correlations For Tests 3 and 4

For both Validity Test 3 and Validity Test 4, we also computed correlations between AutoIC and human-scored IC for each session (Test 3) and each person (Test 4). It is important to note that these are not direct validity comparisons on the same paragraphs (as was the reported average $r = .46$ for Conway et al., 2014), because AutoIC scored all the materials involved and human-scored IC only a fraction (15.9% in Test 3, 3.7% in Test 4) of those materials. As a result of this imprecision, it is unsurprising that the correlations were smaller than those for direct validity tests, and were especially so for the test that only had 3.7% overlap ($r = .28$ for Test 3, $r = .06$ for Test 4). The primary focus of Tests 3 and 4 was to provide replication of the overall key effects, which in the main have higher power in both hand-scored and AutoIC tests than low-power and imprecise correlations. However, to further investigate, for Validity Test 3 for which such data was available, we also computed “like to like” correlations between AutoIC and human-scored IC on the exact same corpus of materials. This correlation ($n = 1100$) was similar to the average reported in the original validity study ($r = .47$; see Table 1). Further – and consistent with Houck et al., (2014) – when these data are aggregated to the session unit of analyses used in Conway et al (2017)’s smoking study, this correlation between AutoIC and human-scored IC was even greater ($r = .70$). These additional analyses suggest (1) that to the degree that data are available that allow for direct validity correlations with human-scored materials, these data show similar levels of validity correlations as in prior work. The further suggest that (2) as in prior work, the larger the “unit” that the comparable data is aggregated at, the higher the correlation between human-scored IC and AutoIC.
Description of Method and Samples for Validation Test 5

For Validation Test 5, we performed a mini meta-analysis (see Goh, Hall, & Rosenthal, 2016) on 11 samples of data collected and scored for AutoIC by the authors. These data purposefully have very little overlap with the original meta-analysis. This is for two reasons: (1) we did not have access to much of the data from the original 35 studies to re-score using AutoIC, and (2) there is actually more theoretical value in replication on a different sample because it uses almost entirely new materials (see Crandall & Sherman, 2016). Thus, this validation test provides added value above and beyond Tests 3 and 4.

Although the present mini meta-analysis is comprised of 11 samples, it nevertheless contains more persons, documents, and paragraphs scored than Houck and Conway’s (2019) meta-analysis of 35 studies. Based on a strategic model and Houck and Conway’s (2019) meta-analysis, we expected that public political samples would show a significant negative relationship between conservatism and complexity; but no such effect would emerge for private citizen samples.

The selection pool was comprised of data collected in projects affiliated with the labs of the authors. From this potential sample, we followed the same inclusion criteria as used in Houck & Conway (2019), except we exclusively incorporated AutoIC studies (and not hand-scored IC). The first author identified the set of studies, and the third author verified that each study met criteria for both political ideology and complexity (see Houck & Conway, 2019 for a full description of criteria for political ideology). This method revealed 11 samples that met our criteria. It is important to note that, even for a small number of samples, mini meta-analyses can be an important and useful tool for researchers (see Goh et al., 2016, for discussion).
The authors then independently coded each sample for whether it was a public political official or private citizen sample. Interrater agreement was 100%. This led to 3 public samples and 8 private samples. Two samples (one public, one private) have overlap with data reported in Houck and Conway (2019): The U.S. presidents’ SOTU speeches, and one of the private samples of political and social topics. However, this overlap is minimal overall, and thus, the majority of the actual scored material is new in this case.

Following identical procedures as Houck & Conway (2019), we used the correlation coefficient $r$ as the effect size estimate between political ideology and AutoIC, and subsequently made appropriate transformations to Fisher’s $z$ (via Lipsey & Wilson, 2001). Transformed effect sizes were used to compute inverse variance weights (Wilson, 2005). Analytic tests using Wilson’s 2005 software macros (also see Lipsey & Wilson, 2001), generated weighted mean effect sizes, confidence intervals, significance levels, and metrics of homogeneity. A random-effects model was used for all analyses in accordance with recommended meta-analytic procedures (Anderson & Bushman, 2001; Burnette et al., 2014; Lipsey & Wilson, 2001; McNamara, Hambrick, & Oswald, 2014; see Houck & Conway, 2019). Following Houck and Conway (2019), we treated different topics completed by the same person in private samples separately. However, we also performed analyses collapsing across topics. When collapsing across topics for private samples (private samples $k = 4$ instead of 8), results were identical to those reported in the main text: Significantly lower complexity for conservatives in public samples ($r = -.12, p = .041, 95\% CI [-.23, -.00]$), but no corresponding effect in private samples ($r = -.03, p = .279, 95\% CI [-.08, +.02]$). Thus, the reported effect is not a result of our strategy of treating private citizens’ topics written by the same person as separate samples.

Below, we describe the samples in more detail.
Sample 1

Sample 1 was drawn from an existing study of U.S. State of the Union (SOTU) speeches, where the first four SOTU speeches were scored for AutoIC for most presidents (Conway & Zubrod, 2020). All material not from the primary speaker was removed prior to scoring (speaker $n = 40$; speech $n = 147$). Ideology was scored using the liberal/conservative ideology historian ratings reported in Thoemmes and Conway (2007).

Sample 2

Sample 2 was drawn from an existing study of U.S. Presidential Debate scoring, where each Presidential debate transcript was scored for AutoIC (Conway & Zubrod, 2020). All material not from the primary speaker was removed prior to scoring (speaker $n = 24$; debate $n = 62$). Ideology was coded as conservative or liberal party identification (“Republican = conservative” and “Democrat” = liberal); all independents were dropped for this analysis.

Sample 3

Sample 3 was drawn from an existing study that scored all 2016 Presidential Primaries debates for AutoIC (Conway et al., 2020). All material not from the speaker in question was removed prior to scoring (speaker $n = 21$; debate $n = 142$). Ideology was coded as conservative or liberal party identification (“Republican = conservative” and “Democrat” = liberal).

Samples 4 and 5

Samples 4 and 5 were drawn from an existing study evaluating the complexity of Right-Wing Authoritarians and Left-Wing Authoritarians (Conway et al., 2020). In addition to other measurements, participants ($n = 202$) in the study completed a two-item self-identification political ideology measurement (items anchored by “liberal/conservative” and “democrat/republican”) that is standard for ideology research (e.g., Jost et al., 2008).
Participants further completed two open-ended stems which were both scored for AutoIC: “Current Republican Party Leadership” (Sample 4) and “Current Democratic Party Leadership” (Sample 5).

**Samples 6 and 7**

Samples 6 and 7 were drawn from an existing study evaluating the complexity of smoking- and health-related attitudes (Conway et al., 2020). In addition to other measurements, participants ($n = 4764$) in the study completed a two-item self-identification political ideology measurement (items anchored by “liberal/conservative” and “democrat/republican”) that is standard for ideology research (e.g., Jost et al., 2008). Participants further completed two open-ended responses to opinion stems which were both scored for AutoIC: “Democrats and Republicans can work together to solve our nations problems” (Sample 6) and one of several potential questions (randomly assigned) pertaining to smoking attitudes and motivations (Sample 7).

**Sample 8**

Sample 8 was drawn from a previously-published study on reaction times and complexity (Conway et al., 2008) that was later scored for AutoIC (Conway et al., 2014). In addition to other measurements, participants ($n = 325$) in the study completed a two-item self-identification political ideology measurement (items anchored by “liberal/conservative” and “democrat/republican”) that is standard for ideology research (e.g., Jost et al., 2008). Participants further randomly assigned to complete one of several open-ended stems that dealt with both political (e.g., “Easy access to birth control”) or personal preference (e.g., “Roller coasters”) topics. All responses were subsequently scored for AutoIC.
Samples 9-11

Samples 9, 10, and 11 were drawn from an existing study evaluating the complexity of responses to political scenarios for liberals and conservatives (Conway et al., 2020). In addition to other measurements, participants ($n = 501$) in the study completed a one-item self-identification political ideology measurement (item anchored by “very liberal/very conservative”) that is standard for ideology research (e.g., Jost et al., 2008). Participants read either two or three scenarios and then gave open-ended responses to opinion stems about those scenarios. 501 participants completed two of the scenarios; 249 participants completed a third scenarios. The first scenario (Sample 9) involved participant responses to a proposed law concerning prayer in schools. The second scenario (Sample 10) involved the rights of political protestors. The third scenarios (Sample 11) involved a military officer questioning and order from the President. All open-ended responses were scored for AutoIC.
Integrative Complexity and AutoIC Supplement

At a measurement level, integrative complexity is scored on a 1-7 scale. Specifically, the absence of differentiation reflects cognitive simplicity, resulting in scores of 1. For example, the statement “World War II helped free the world from tyranny” contains only 1 dimension (freedom from tyranny) associated with one target (World War II) and would thus earn a score of 1. Differentiation is a recognition of different perspectives on an issue; communication with differentiation (but no integration) is given a score of 3. For example, the statement “World War II helped free the world from tyranny, but at the same time, it cost so many innocent lives” involves the recognition of two dimensions (freedom, loss of life) and would thus earn a score of 3. Scores above 3 occur when connections are drawn between differentiated dimensions. For example, consider the statement “World War II helped free the world from tyranny, but it cost so many lives. However, the loss of life often is proportionally related to freedom from tyranny, and thus the two things are intertwined.” That statement identifies two separate dimensions (tyranny, loss of life) and further discusses how they are integrated proportionally. As a result, it would earn a score of 5. Scores higher than 5 occur rarely and are the result of hierarchical analyses that involve different integrated dimensions.

AutoIC scores documents by breaking them down into paragraphs of equal length (or by scoring at a paragraph-by-paragraph level if researchers choose that option). In scoring, AutoIC looks within each paragraph for whole words/phrases from the AutoIC dictionary. It then applies an algorithm to phrases that match the dictionary. This algorithm assigns scores based on the probable complexity associated with each word/phrase. Below, we illustrate by exploring some sample paragraphs scored by a human expert and AutoIC.
Paragraph 1: One of those memories is sitting on my grandfather’s shoulders and watching the Apollo astronauts come ashore in Hawaii. People cheered and waved small flags, and my grandfather explained with pride and assurance how we Americans could accomplish anything we set our minds to do. I lived overseas for a time as a child, and I remember listening to my mother reading me the first lines of the Declaration of Independence and explaining how its ideas applied to every American. She told me that those words protected us from the brutal injustices we witnessed other people suffering during those years abroad.

The above paragraph was scored a 1 by an expert human scorer. It contains no differentiation, rather just a long list of descriptions. It was also scored a 1 by AutoIC.

Paragraph 2: As with most Americans, patriotism starts for me as a gut instinct, a loyalty and love for my country that’s rooted in my earliest memories. It’s not just the recitations of the Pledge of Allegiance, the Thanksgiving pageants at school, or the fireworks on the Fourth of July, but how the American ideal wove its way throughout the lessons my family taught me.

The above paragraph was scored a 3 by an expert coder because of the clear differentiation between the more superficial reasons (recitations, etc.) associated with patriotism and the weaving of the American ideals. AutoIC scored it a 2.5 due to the differentiating phrase “not just the.”

Paragraph 3: In return, the gift we can give back to our country is a patriotism that requires us to be good citizens in public office or in the community spaces where government is
absent. We should, by all means, argue with each other, as did Adams and Jefferson, about the policies of government and the history we hope to make tomorrow. But it should be an argument among friends, who agree more than they disagree, each of us united in a cause larger than our individual interests, honestly debating the best means to serve that cause, and intent on finding some common ground upon which to overcome together the many challenges before us. To love one’s country is to love one’s countrymen. And if we are to replicate the spirit of our founding age, if we are to be genuine patriots, we must remember also that we are patriots because we love the countrymen we will never know, who will be born after we are gone.

The above paragraph was scored a 5 by an expert coder because of the clear differentiation between the various parts of the “argument,” plus the clear attempt at integration to find common ground. AutoIC scored the paragraph a 4.75 due to (a) multiple markers of differentiation, including “many challenges,” “and because,” and “also.” (b) Further, it was scored as having multiple examples of integration language, including the phrases “common ground,” “argument among friends,” “in return,” and “give back.”
Additional References


