

Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly

DANIEL M. OPPENHEIMER*

Princeton University, USA

SUMMARY

Most texts on writing style encourage authors to avoid overly-complex words. However, a majority of undergraduates admit to deliberately increasing the complexity of their vocabulary so as to give the impression of intelligence. This paper explores the extent to which this strategy is effective. Experiments 1–3 manipulate complexity of texts and find a negative relationship between complexity and judged intelligence. This relationship held regardless of the quality of the original essay, and irrespective of the participants' prior expectations of essay quality. The negative impact of complexity was mediated by processing fluency. Experiment 4 directly manipulated fluency and found that texts in hard to read fonts are judged to come from less intelligent authors. Experiment 5 investigated discounting of fluency. When obvious causes for low fluency exist that are not relevant to the judgement at hand, people reduce their reliance on fluency as a cue; in fact, in an effort not to be influenced by the irrelevant source of fluency, they over-compensate and are biased in the opposite direction. Implications and applications are discussed. Copyright © 2005 John Wiley & Sons, Ltd.

When it comes to writing, most experts agree that clarity, simplicity and parsimony are ideals that authors should strive for. In their classic manual of style, Strunk and White (1979) encourage authors to 'omit needless words.' Daryl Bem's (1995) guidelines for submission to *Psychological Bulletin* advise, 'the first step towards clarity is writing simply.' Even the APA publication manual (1996) recommends, 'direct, declarative sentences with simple common words are usually best.'

However, most of us can likely recall having read papers, either by colleagues or students, in which the author appears to be deliberately using overly complex words. Experience suggests that the experts' advice contrasts with prevailing wisdom on how to sound more intelligent as a writer. In fact, when 110 Stanford undergraduates were polled about their writing habits, most of them admitted that they had made their writing more complex in order to appear smarter. For example, when asked, 'Have you ever changed the words in an academic essay to make the essay sound more valid or intelligent by using complicated language?' 86.4% of the sample admitted to having done so. Nearly two-thirds answered yes to the question, 'When you write an essay, do you turn to the thesaurus to choose words that are more complex to give the impression that the content is more valid or intelligent?'

*Correspondence to: D. M. Oppenheimer, Department of Psychology, Princeton University, Green Hall Room 2-S-8, Princeton, NJ 08540, USA. E-mail: doppenhe@princeton.edu

There are many plausible reasons that the use of million-dollar words would lead readers to believe that an author is smart. Intelligence and large vocabularies are positively correlated (Spearman, 1904). Therefore, by displaying a large vocabulary, one may be providing cues that he or she is intelligent as well. Secondly, writers are assumed to be conforming to the Gricean maxim of manner, 'avoid obscurity of expression' (Grice, 1975). If authors are believed to be writing as simply as possible, but a text is nonetheless complex, a reader might believe that the ideas expressed in that text are also complex, defying all attempts to simplify the language. Further, individuals forced to struggle through a complex text might experience dissonance if they believe that the ideas being conveyed are simple (Festinger, 1957). Thus, individuals might be motivated to perceive a difficult text as being more worthwhile, thereby justifying the effort of processing.

Indeed, there is some evidence that complex vocabulary can be indicative of a more intelligent author. For example, Pennebaker and King (1999) have shown that the percentage of long words used in class assignments positively correlates with SAT scores and exam grades on both multiple choice and essay tests. However it is difficult to draw conclusions about the effectiveness of a strategy of complexity from this data. The study did not look at how readers of the texts containing the long words perceived the authors' intelligence. Thus, it is possible that although students using complex vocabularies are objectively very knowledgeable, they might nonetheless be perceived as being less so.

Why might we believe that the experts might be correct in recommending simplicity in writing? One theory that predicts the effectiveness of straightforward writing is that of processing fluency. Simpler writing is easier to process, and studies have demonstrated that processing fluency is associated with a variety of positive dimensions. Fluency leads to higher judgements of truth (Reber & Schwarz, 1999), confidence (Norwick & Epley, 2002), frequency (Tversky & Kahneman, 1973), fame (Jacoby, Kelley, Brown, & Jasechko, 1989), and even liking (Reber, Winkielman, & Schwarz, 1998). Furthermore, the effects of fluency are strongest when the fluency is discrepant—when the amount of experienced fluency is surprising (Whittlesea & Williams, 2001a, 2001b). As such, it would not be surprising if the lower fluency of overly complex texts caused readers to have negative evaluations of those texts and the associated authors, especially if the complexity was unnecessary and thus surprising readers with the relative disfluency of the text.

Both the experts and prevailing wisdom present plausible views, but which (if either) is correct? The present paper provides an empirical investigation of the strategy of complexity, and finds such a strategy to be unsuccessful. Five studies demonstrate that the loss of fluency due to needless complexity in a text negatively impacts raters' assessments of the text's authors.

EXPERIMENT 1

Experiment 1 aimed to answer several simple questions. First, does increasing the complexity of text succeed in making the author appear more intelligent? Second, to what extent does the success of this strategy depend on the quality of the original, simpler writing? Finally, if the strategy is unsuccessful, is the failure of the strategy due to loss of fluency? To answer these questions, graduate school admission essays were made more complex by substituting some of the original words with their longest applicable thesaurus entries.

While word length is not perfectly interchangeable with sentence complexity—for example, complexity can come from grammatical structure or infrequent words as well—it is a useful proxy. Using length as a manipulation of complexity allows for a simple, easily replicable word replacement algorithm. By keeping content constant and varying the complexity of vocabulary, it was possible to investigate the effectiveness of complexity.

Participants and procedure

Seventy-one Stanford University undergraduates participated to fulfil part of a course requirement. The survey was included in a packet of unrelated one-page questionnaires. Packets were distributed in class, and participants were given a week to complete the entire packet.

Stimuli and design

Six personal statements for admissions to graduate studies in English Literature were downloaded from writing improvement websites. The essays varied greatly both in content and quality of writing. Logical excerpts ranging from 138 to 253 words in length were then taken from each essay. A ‘highly complex’ version of each excerpt was prepared by replacing every noun, verb and adjective with its longest entry in the Microsoft Word 2000 thesaurus. Words that were longer than any thesaurus entry, were not listed in the thesaurus, or for which there was no entry with the same linguistic sense were not replaced. If two entries were of the same length, the replacement was chosen alphabetically. When necessary, minor modifications were made to the essay to maintain the grammatical structure of a sentence (e.g. replacing ‘an’ with ‘a’ for replacement words beginning with consonants). A ‘moderately complex’ version of each excerpt was created using the same algorithm as above, except replacing only every third applicable word. Examples of the stimuli can be found in the appendix.

Each participant received only one excerpt. Participants were informed that the excerpt came from a personal statement for graduate study in the Stanford English department. They were instructed to read the passage, decide whether or not to accept the applicant, and rate their confidence in their decision on a 7-point scale.¹ They were then asked how difficult the passage was to understand, also on a seven-point scale.

Results

The data of one participant was discarded due to an illegible answer. Analysis of the manipulation check showed that more complex texts were more difficult to read. ($x = 2.9$, 4.0 and 4.3 for simple, moderately complex and highly complex, respectively). These differences were reliable, $F(2, 68) = 4.46$, $p < 0.05$, Cohen’s $f = 0.18$. For other analyses, acceptance ratings (+1 for accept, -1 for reject) were multiplied by confidence ratings to create a -7 to 7 scale of admission confidence. Level of complexity had a reliable influence on admission confidence ratings, $F(2, 70) = 2.46$, $p < 0.05$, Cohen’s $f = 0.12$.

¹With the exception of the dichotomous admissions decision, all dependent measures reported in this paper are seven point scales ranging from 1 = ‘not at all’ to 7 = ‘very’.

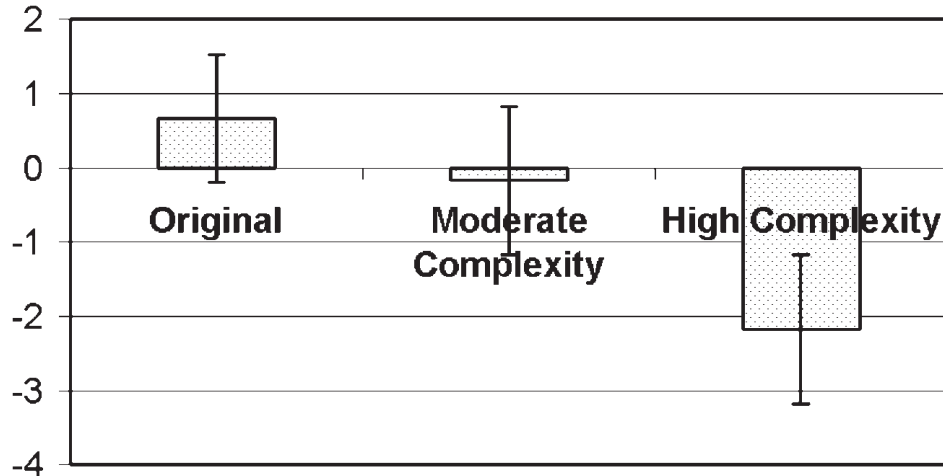


Figure 1. Acceptance ratings (on a -7 to 7 scale) for each level of complexity

Highly complex essays (mean = -2.1) were rated more negatively than moderately complex essays (mean = -0.17), which in turn were rated more negatively than the original essays ($x = 0.67$).² These differences are summarized in Figure 1. Additionally, the excerpts reliably varied in quality; average admissions confidence ratings ranged from -3.1 to 1.8 $F(5, 70) = 2.2$, $p < 0.05$, Cohen's $f = 0.12$. However, there was no reliable interaction between the quality of the initial excerpt and the level of complexity $F(10, 70) = 1.4$, $p > 0.10$, Cohen's $f = 0.07$.

To determine if the negative influence of complexity on admissions ratings was due to differences in fluency, a mediation analysis was run using difficulty of comprehension as a mediator. Level of complexity was reliably correlated with acceptance ratings, $r(69) = -0.24$, $p < 0.05$ and difficulty of comprehension $r(69) = 0.32$, $p < 0.05$. However, when controlling for difficulty of comprehension, the relationship between complexity and acceptance was drastically reduced $r(69) = -0.14$, $p > 0.1$, while controlling for complexity did not remove the relationship between difficulty and acceptance $r(69) = -0.25$, $p < 0.05$. A Sobel test demonstrated this mediation to be reliable, $z = 2.1$, $p < 0.05$. These results are summarized in Figure 2.

Discussion

The results of Experiment 1 suggest that contrary to prevailing wisdom, increasing the complexity of a text does not cause an essay's author to seem more intelligent. In fact, the opposite appears to be true. Complex texts were less likely than clear texts to lead to acceptance decisions in a simulated admissions review. Simple texts were given higher ratings than moderately complex texts, which were, in turn, given better ratings than highly complex texts. Additionally, this trend was found regardless of the quality of the original essay. Complexity neither disguised the shortcomings of poor essays, nor enhanced the appeal of high-quality essays. The mediation analysis suggests that the

²Post-hoc analysis revealed that the 'moderate complexity' condition was not reliably different from either the 'highly complexity' or control conditions.

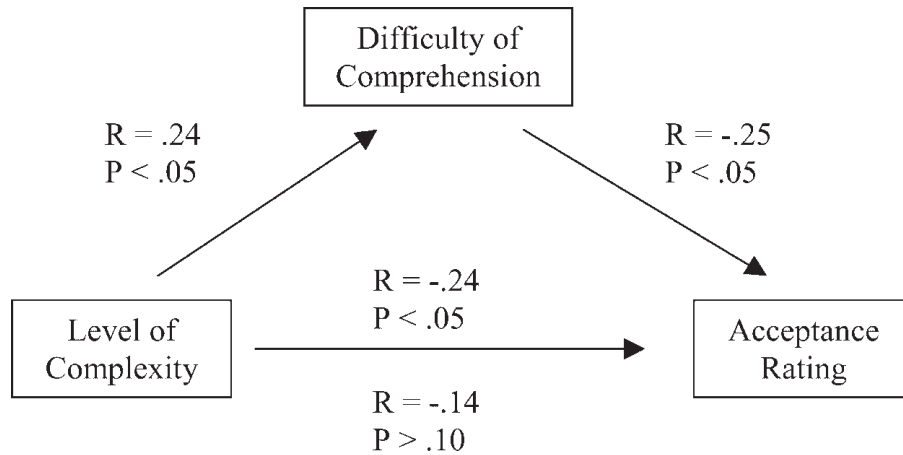


Figure 2. Mediation analysis in Experiment 1

reason that simple texts are viewed more positively than complex texts was due to fluency. Complex texts are difficult to read, which in turn leads to lower ratings.

Even though Experiment 1 is suggestive, there are several problems that need to be resolved before any conclusions can be drawn. First, it is possible that the reason that complexity was unsuccessful was that words were misused. In an effort to prevent experimenter biases from influencing the data, the word replacement process was algorithmic, and left little room for human judgement. Although only synonyms of the appropriate linguistic sense were included, and grammatical editing took place, it is nonetheless possible that some of the replacement words were used slightly out of context, or led to awkward sounding sentences.

Secondly, the domain of college application essays may lead to biases against the strategy of complexity. Participants likely are aware of the widespread use of the strategy—especially in admissions essays—and may be actively discounting the use of complex words. Finally, it could be the case that complexity is differentially successful as a strategy depending upon a reader's prior expectation of the author's intelligence. In Experiment 1, the readers had no reason to think that the authors were particularly intelligent; maybe if the readers had believed the authors to be brilliant at the outset of the experiment, the presence of complex vocabulary would have reinforced such a belief and led to higher ratings.

As such, a second experiment was run to control for the confounds in Experiment 1 and investigate the impact of prior beliefs.

EXPERIMENT 2

If actively replacing words in an essay may impair the quality of the text, then to test the effects of complex words we need a more natural set of stimuli. Therefore, for Experiment 2 it was necessary to find two essays of identical content, but using different vocabulary, in which the experimenters did not influence word selection. Many texts in foreign languages have multiple translations, which conform to the original meaning of the text, but use different words and grammatical construction. This provides the perfect domain for testing whether complex phrasing and vocabulary hurts perceptions of a text.

Participants and procedure

Thirty-nine Stanford University undergraduates participated to fulfil part of a course requirement. The survey was included in a packet of unrelated one-page questionnaires. Packets were distributed in class and participants were given a week to complete the entire packet.

Stimuli and design

Translations of the first paragraph of Rene Descartes *Meditation IV* were sought until two renditions of comparable word counts, but contrasting complexity were found. Heffernan's (1990) 98-word translation was judged by two independent raters to be considerably more complex than Tweyman's (1993) 82-word version. The exact stimuli can be found in the appendix.

Each translation was read by half of the participants. Additionally, to manipulate prior expectations of author intelligence, half of the participants were told that the passage came from Descartes, while the rest were told that it came from an anonymous author.³ Participants were instructed to read the passage and rate the intelligence of the author on a 7-point scale. They were then asked how difficult the passage was to understand, also on a 7-point scale; this question served both as a measure of fluency, and as a manipulation check to verify the difference in complexity of the translations.

Results

Analysis of the manipulation check showed that the Heffernan (1990) translation (mean complexity rating = 5.4) was indeed perceived as more complex than the Tweyman (1993) translation (mean complexity rating = 4.5), $t(37) = 1.77$, $p < 0.05$, Cohen's $d = 0.58$. There were reliable main effects for both complexity, $F(1, 39) = 3.65$, $p < 0.05$, Cohen's $f = 0.18$, and prior belief, $F(1, 39) = 17.36$, $p < 0.05$, Cohen's $f = 0.45$; participants who read the simpler translation and attributed it to Descartes rated the author as more intelligent (mean = 6.5) than those reading the complex translation attributed to Descartes (mean = 5.6). Those who were given no source for the passage also rated the author as more intelligent in the simple version (mean = 4.7) than the complex version (mean = 4.0). However, there was no reliable interaction between prior belief and level of complexity, $F(1, 39) = 0.08$, $p > 0.10$, Cohen's $f = 0.00$. The results are summarized in Figure 3.

To determine if the negative influence of complexity on intelligence ratings was due to differences in fluency, a mediation analysis was run using difficulty of comprehension as a mediator.⁴ Complexity was reliably correlated negatively with intelligence ratings, $r(37) = -0.30$, $p < 0.05$, and positively with difficulty of comprehension, $r(37) = 0.33$, $p < 0.05$. However, when controlling for difficulty of comprehension, the relationship between complexity and intelligence ratings was reduced, although still marginally significant, $r(37) = -0.24$, $0.05 > p < 0.1$, while controlling for complexity did not remove the relationship between difficulty and intelligence ratings, $r(37) = -0.28$,

³Participants would all know who Descartes was, as they had all read his work (although not *Meditation IV*) in the introduction to humanities class that all Stanford students are required to take.

⁴Level of prior belief was statistically controlled for in all correlations reported here.

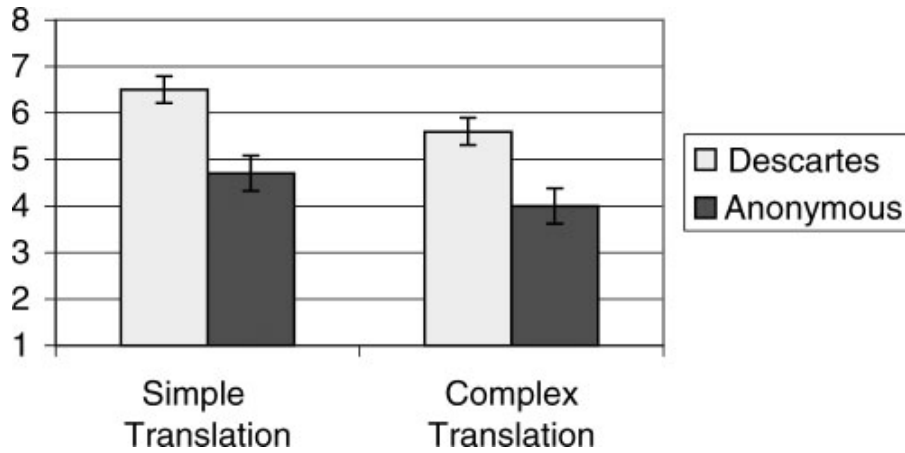


Figure 3. Intelligence ratings of the authors of two different translations of Descartes *Meditation IV*, when attributed either to Descartes or to an anonymous author

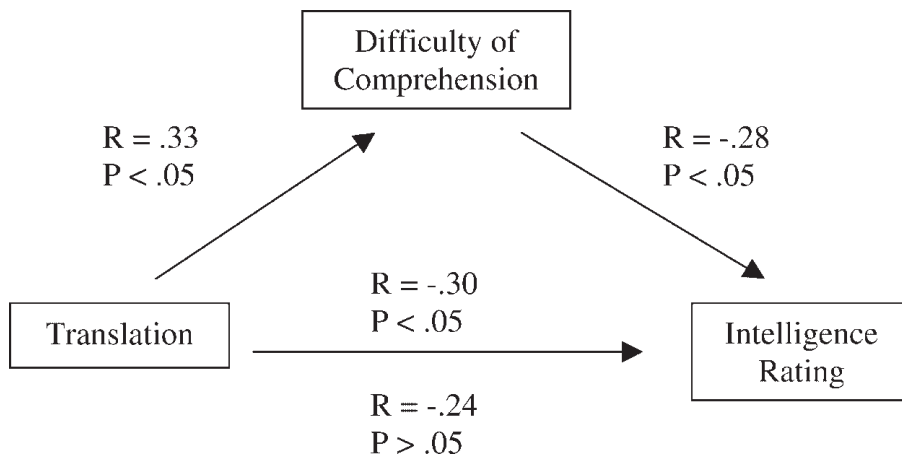


Figure 4. Mediation analysis in Experiment 2

$p < 0.05$. While these results are in the right direction and suggest a mediation effect, they do not achieve statistical significance when analysed by a Sobel test, $z = 1.2, p > 0.05$. The results are summarized in Figure 4.

Discussion

The results of Experiment 2 support those of Experiment 1. Once again, complexity negatively influenced raters’ assessments of texts. This relationship was found regardless of the raters’ prior expectations of the author’s intelligence. While the data suggest that the process may be mediated by fluency, the failure to reach statistical significance means that it is difficult to draw strong conclusions. However, in light of the fact that the mediation analysis *was* reliable in Experiment 1, and was in the predicted direction for Experiment 2, normatively one should have increased confidence in the reliability of the effect (Tversky & Kahneman, 1971). This is especially true in light of the fact that Sobel tests have been shown to be overly conservative estimators of statistical significance (Mackinnon, Lockwood, Hoffman, West, & Sheets, 2002).

However, aside from the mediation analysis, there are other challenges in interpreting this experiment. Some translators are better than others. A less accomplished translator might create a less fluent text for reasons completely unrelated to word complexity. It seems possible that the reason that the more complex text was judged to have come from a less intelligent author was simply because the translation was not as skillful.

Thus, results from the first two experiments could be due to the fact that the complex essays were in actuality worse papers. As such, it was important to run a third study to try and ensure that the lower ratings are due to the use of complex vocabulary instead of inferior quality papers.

EXPERIMENT 3

The word replacement paradigm used in Experiment 1 was problematic because using an algorithmic approach to word replacement leads to the possibility of including imprecise synonyms, impairing flow and generally making the essay less coherent. If it were indeed the case that algorithmic word replacement leads to poorer essays, then one would expect that the process should also harm an essay modified to use simpler vocabulary. However the fluency account leads to the opposite prediction; less complex essays should be rated as coming from more intelligent authors. To test these contrasting predictions Experiment 3 used the same procedure as Experiment 1 but systematically simplified text.

Participants and procedure

Thirty-five Stanford University undergraduates participated to fulfil part of a course requirement. Surveys were included in a packet of unrelated one-page questionnaires that were filled out in a one-hour lab session. An additional 50 Stanford University undergraduates were recruited outside of dining halls and filled out only the relevant survey.

Stimuli and design

Twenty-five randomly chosen dissertation abstracts from the Stanford University sociology department were examined, and the abstract with the highest proportion of words of nine letters or longer was chosen (Chang, 1993). The first two paragraphs (144 words) were taken from the abstract. A 'simplified' version of each the excerpt was prepared by replacing every word of nine or more letters with its second shortest entry in the Microsoft Word 2000 thesaurus. Words that were shorter than any thesaurus entry, were not listed in the thesaurus, or for which there was no entry with the same linguistic sense were not replaced. If two entries were of the same length, the replacement was chosen alphabetically. When necessary, minor modifications were made to the essay to maintain the grammatical structure of a sentence (e.g. replacing 'an' with 'a' for replacement words beginning with consonants). Excerpts from the stimuli can be found in the appendix.

Participants were informed that the excerpt came from a sociology dissertation abstract. Participants were instructed to read the passage and rate the intelligence of the author on a 7-point scale. They were then asked how difficult the passage was to understand, also on a 7-point scale.

Results

Analysis of the manipulation check showed that the ‘simplified’ version was indeed perceived as less complex (mean complexity rating = 4.9) than the original excerpt (mean complexity rating = 5.6), $t(83) = 2.327$, $p < 0.05$, Cohen’s $d = 0.53$. There was also a reliable effect of complexity on intelligence judgements; participants who read the ‘simplified’ version rated the author as more intelligent (mean = 4.80) than those reading the original version (mean = 4.26), $t(83) = 1.988$, $p < 0.05$, Cohen’s $d = 0.44$.

To determine if the negative influence of complexity on intelligence judgements was due to differences in fluency, a mediation analysis was run using difficulty of comprehension as a mediator. Complexity was reliably correlated with intelligence ratings, $r(85) = -0.213$, $p < 0.05$, and difficulty of comprehension, $r(85) = -0.247$, $p < 0.05$. However, when controlling for difficulty of comprehension, the relationship between complexity and intelligence ratings was reduced, although still marginally significant, $r(85) = -0.196$, $0.05 > p < 0.1$. While these results are in the right direction and suggest a mediation effect, they do not achieve statistical significance when analyzed by a Sobel test, $z = 0.75$, $p > 0.05$.

Discussion

The results of Experiment 3 further support the notion that the use of overly complex words leads to lower evaluations of a text’s author. While in Experiment 1 it could be argued that the replacement of words leads to stilted sounding text, in Experiment 3 the word-replacement condition actually increased judgements of intelligence. Further, given the fact that the replacement process was algorithmic, it seems unlikely that the improvements in the essays could be due to editing or experimenter bias. It is the use of overly complex words—not the word replacement process—that leads to decreased ratings of intelligence.

Additionally, in all three experiments the result appears to be at least partially mediated by fluency. In all experiments the data conforms to the pattern that one would expect if fluency were a mediator, and in Experiment 1 this pattern is demonstrated to be reliable. This fits well into Kahneman and Frederick’s (2002) notion of attribute substitution; rating a person’s intelligence or suitability for graduate admission is difficult, so people might use fluency as a proxy for these judgements. However, it is difficult to conclude that fluency is necessarily responsible for the effect because there was no direct manipulation of fluency in the first three experiments. Further, the lack of statistical reliability in the mediation analyses from Experiments 2 and 3 led to questions about whether the lowered evaluations of the complex text were due to fluency at all. Thus, it seems worthwhile to further explore the mechanism behind why added complexity lowers ratings of intelligence.

EXPERIMENT 4

If the fluency hypothesis is correct, then any manipulation that substantially reduces fluency should also reduce intelligence ratings. One method that has proven to be effective in reducing fluency is presenting the text in a font that is difficult to read (Norwick &

Epley, 2002). By manipulating font, it was possible to examine whether fluency can influence intelligence ratings directly, or whether there was an unmeasured variable driving the mediation effects in Experiments 1–3.

Participants and procedure

Fifty-one Stanford University undergraduates participated to fulfil part of a course requirement. The survey was included in a packet of unrelated one-page questionnaires. Packets were distributed in class, and participants were given a week to complete the entire packet.

Stimuli and design

The unedited version of the highest quality essay from Experiment 1 (163 words) was used. A ‘non-fluent’ version of the excerpt was prepared by converting the document into italicized ‘Juice ITC’ font. The original version was in normal ‘Times New Roman’ font. Both versions used 12-point typeset. For an illustration of each of fonts, please see Figure 5.

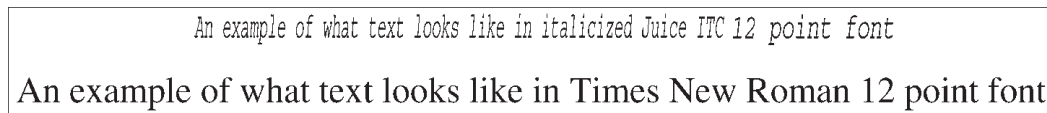


Figure 5. Illustrations of the fonts in both the fluent and non-fluent versions of the questionnaire

Each participant received only one excerpt. Participants were informed that the excerpt came from a personal statement for graduate study in the Stanford English Department. They were instructed to read the passage, and rate the author’s intelligence on a 7-point scale. To prevent participants from believing that the author of the text had chosen that font (as font selection could be cue about intelligence) the instructions and rating scales were also written in the corresponding font. Thus, participants would attribute font selection to the experimenter instead of the text’s author.

Results

Post-experimental interviews of randomly selected participants ($n = 5$) confirmed that participants attributed the font selection to the experimenter rather than to the author of the essay. There was a reliable effect of font on intelligence judgements; participants who read the ‘non-fluent’ version rated the author as less intelligent (mean = 4.04) than those reading the original version (mean = 4.50), $t(49) = 1.69$, $p < 0.05$ one-tailed, Cohen’s $d = 0.48$.

Discussion

Experiment 4 directly manipulated fluency, and found that fluency impacted intelligence ratings. When texts were written in a font that was difficult to read, the author of the text was judged to be less intelligent. Taken in conjunction with the mediation analyses in

Experiments 1–3, this strongly suggests that complex vocabulary makes texts harder to read, which in turn lowers judgements of an author's intelligence.

If, as Experiments 1–4 suggest, fluency is the driving factor behind these effects, then one ought to be able to reverse the direction of the effect by making people aware that the source of the low fluency is irrelevant to judgement. People tend to attribute events to a single cause, rather than multiple causes (Einhorn & Hogarth, 1986; Kelley, 1973). Thus, when one cause is known to have occurred, people think that other causes are less likely to have also occurred. This phenomenon applies to the metacognitive experience of fluency (Oppenheimer, 2004; Schwarz, 2004; Whittlesea & Williams, 1998). When obvious causes for low fluency exist that are not relevant to the judgement that is being made, people reduce their reliance on fluency as a cue; in fact, in an effort not to be influenced by the irrelevant source of fluency, they overcompensate and are biased in the opposite direction (see Wilson & Brekke, 1994 for a review of overcompensation effects).

For example, Oppenheimer (2004) asked people to make judgements about surname frequency, a task for which people typically use fluency as a cue (Tversky & Kahneman, 1973). In a series of experiments, he showed that in the presence of obvious causes for fluency that had no bearing on frequency—such as personal relevance, or a famous individual associated with that name—people no longer used fluency in making their judgement. In fact, they tended to rate the fluent name as less frequent rather than more frequent when a salient cause for fluency was available.

Spontaneous discounting of fluency suggests that conscious awareness of the source of low fluency should undermine the effectiveness of the fluency manipulation. In fact, if there is an obvious cause for lack of fluency the trends might actually reverse as people overcompensate in their attempt not to be influenced by fluency. Experiment 5 investigates this possibility.

EXPERIMENT 5

One method for lowering fluency and making the source of the decreased fluency obvious, is the 'low toner' paradigm (Oppenheimer & Frank, under review). Documents printed from a printer that is low in toner are hard to read because the text is not as dark on the page as usual, and the text has streaks running through it. However, the cause of the lack of fluency is immediately obvious to anybody who has ever observed a low toner document. Because the reason for the low fluency will be obvious to participants, a fluency account would predict that people would discount their lack of fluency. In an effort not to be influenced by the irrelevant fluency information, people are likely to overcompensate, and have their judgements skewed in the other direction (Oppenheimer, 2004).

Method

Participants and procedure

Twenty-seven Stanford University undergraduates participated to fulfil part of a course requirement. The survey was included in a packet of unrelated one-page questionnaires. Packets were distributed in class, and participants were given a week to complete the entire packet.

Stimuli and design

The unedited version of a randomly chosen essay from Experiment 1 was used. Both conditions were prepared using standard 12 point 'Times New Roman' font. The 'non-fluent' version of the excerpt was created by waiting until the departmental printer was low on toner, and printing the surveys out while the toner was low. For a scanned image of the stimuli, please see Figure 6.

Each participant received the excerpt either in normal or low-toner font. Participants were informed that the excerpt came from a personal statement for graduate study in the Stanford English Department. They were instructed to read the passage, decide whether or not to accept the applicant, and rate their confidence in their decision on a 7-point scale (as in Experiment 1). They were also asked to rate the author's intelligence on a 7-point scale (as in Experiments 2 and 3).

Personal Statements for Graduate Studies in the Stanford University English Department: *Please read the following personal statement, and answer the questions below.*

Through six years of teaching experience in the public schools, nothing has proved more challenging and intriguing to me than the teaching of writing. My experience in writing instruction has led me to two principal areas which I desire to research further in my graduate work: computers in the composing process and writing across the curriculum. Over the past three years I have become especially interested in the developing field of composing with computers. I have used my English classes as informal laboratories to observe different stages of writing processes on computers. My current experimentation is in the use of computers as tools for response and revision via a distance network. In my graduate study, I would like to further explore the range of possibilities of computers and networks assisting in revision processes. I am also interested in the design of computer writing centers and would like to study how these environments both change our approaches and aid us in our writing tasks.

Personal Statements for Graduate Studies in the Stanford University English Department: *Please read the following personal statement, and answer the questions below.*

Through six years of teaching experience in the public schools, nothing has proved more challenging and intriguing to me than the teaching of writing. My experience in writing instruction has led me to two principal areas which I desire to research further in my graduate work: computers in the composing process and writing across the curriculum. Over the past three years I have become especially interested in the developing field of composing with computers. I have used my English classes as informal laboratories to observe different stages of writing processes on computers. My current experimentation is in the use of computers as tools for response and revision via a distance network. In my graduate study, I would like to further explore the range of possibilities of computers and networks assisting in revision processes. I am also interested in the design of computer writing centers and would like to study how these environments both change our approaches and aid us in our writing tasks.

Figure 6. Scanned images of the low toner version of the excerpt, and the original excerpt that were used in Experiment 5

Results

As in Experiment 1, acceptance ratings (+1 for accept, -1 for reject) were multiplied by confidence ratings to create a -7 to 7 scale of admission confidence. As predicted, participants in the low toner condition were more likely to recommend acceptance for the applicant (mean = 2.0) than those in the normal font condition (mean = -1.8). This difference was reliable, $t(25) = 2.15$, $p < 0.05$, Cohen's $d = 0.86$. Additionally, participants in the low toner conditions reliably rated the author as more intelligent (mean = 5.0) than those in the normal condition (mean = 4.0), $t(25) = 2.72$, $p < 0.05$, Cohen's $d = 1.09$.

Discussion

As predicted by the fluency account, when an obvious source for the lack of fluency is present, people discount that lack of fluency when making their judgement. They do so to such an extent that they end up biasing their judgement in the opposite direction! This trend can not be explained by unpleasant mood lowering ratings across the board. Instead, the effect seems to be constrained by the manner in which fluency is processed; when there is no obvious source of fluency (Experiment 4) then intelligence judgements are lowered, but in the presence of an obvious source of fluency (Experiment 5) intelligence judgements increase.⁵

GENERAL DISCUSSION

In the first three experiments, the negative consequences of needless complexity were shown in widely disparate domains (personal statements, sociology dissertation abstracts and philosophical essays), across different types of judgements (acceptance decisions and intelligence ratings), and using distinct paradigms (active word replacement and translation differences). The effect was demonstrated regardless of the quality of the original essay or prior beliefs about a text's quality. All in all, the effect is extremely robust: needless complexity leads to negative evaluations.

The results further suggest that this effect is due to lowered processing fluency. Experiment 4 shows that directly reducing fluency through a standard font manipulation (e.g. Norwick & Epley, 2002) leads to lower intelligence judgements. Further, Experiment 5 demonstrated that if the source of reduced fluency becomes obvious, participants will discount their lack of fluency, which reverses the direction of the effect. Mediation analyses in Experiments 1-3 suggest a similar process is occurring with complex vocabulary. However, it is worth noting that although Experiment 1 and Experiment 3 were conceptually very similar, the results of the mediation analyses varied in regard to their reliability. This suggests that while fluency clearly influences intelligence judgements, there are most certainly other factors in play as well. For

⁵One question that arises from this study is what sources are 'obvious' enough to elicit spontaneous discounting. A challenge that arises in answering this question, is that how 'obvious' the source needs to be varies depending on the situation; sources need to be much more salient to elicit discounting when time constraints are imposed, and need be much less so when participants are highly motivated to thoroughly think through their judgements (Oppenheimer & Monin, in prep). Further investigation in this area is clearly important, although well beyond the scope of the current paper.

example, attributions of author intent or expectations about standard complexity in a given domain (e.g. the belief that dissertations should be more complex than admissions essays) might very well play a role as well. While it is undoubtedly worthwhile to identify and empirically investigate these other factors, doing so is beyond the scope of this paper. At this point, however, one can conclude that fluency clearly is at least partially responsible for the effect, and since longer words lower fluency they can have a negative impact on intelligence judgements.

However, one cannot conclude from these results that using long words is always problematic. For one thing, the population tested in these studies is extremely limited. Stanford students are both well educated and motivated; it is possible that this pattern of results was found only because participants were able to understand the complex vocabulary, and made the effort to muddle through to the content beneath. Similarly, one could imagine that experts in a given field (who are more familiar with the jargon) would react differently to simplified essays than novices. For one, the experts would find the jargon a great deal more fluent than non-experts. Additionally, a lack of jargon might be a signal that the author is not an in-group member of the field; this could lead to simplified writing being negatively associated with intelligence. Thus, further research is necessary to determine if these results generalize to the population as a whole.

Another limitation is that these studies exclusively examined written text; it is unclear whether the same effects would apply to oral conversation as well. Finally, there are many times when a long word is appropriate, because it is more precise or concise. These studies primarily investigated the use of needless complexity in writing. When a long word is actually the best word for the occasion, it very well may be that using it will lead to positive appraisals. Indeed, these studies can not rule out the possibility that in some situations judicious use of a thesaurus will improve the quality of writing. A thesaurus could be used to help select the most appropriate word for a given argument such that decreases in fluency are overridden by increases by other positive attributes of a given word substitution.

It is also worth examining potential boundary conditions of the effect. While the present studies have primarily investigated the impact of complexity on intelligence judgements, it seems possible that other dimensions such as liking, sociability or trustworthiness could be impacted as well. Likely, the extent to which other dimensions are impacted will be related to people's naïve theories of how fluency is related to those dimensions (Schwarz, 2004). If people tend to believe that fluency is positively correlated with sociability, then increasing complexity of a text should lead to lower judgements of the author's sociability. Alternatively, if people tend to believe that fluency is negatively correlated with sociability, then fluency would have the opposite effect on judgement. This leads to the intriguing possibility that if participants could somehow be primed to think that disfluent text tended to come from more intelligent authors, one would expect the results from this set of studies to reverse.

However, it seems that people's naïve theories of fluency tend to lead them to negatively associate complexity and intelligence. This has some interesting ramifications. The most straightforward of these is that authors should avoid needless complexity. As reported in the introduction of this paper, a vast majority of Stanford students use a strategy of complexity when writing papers and this is undoubtedly true at campuses and businesses across the country. However, this research shows that such strategies tend to backfire. This finding could be broadly applied to help people improve their writing, and receive more positive evaluations of their work.

Secondly, there are some exciting potential applications that become apparent by examining when people are more likely to complicate their writing. Pennebaker and Lay (2002) have shown that people are more likely to use big words when they are feeling the most insecure. One can imagine that a minority student under stereotype threat (Steele, 1997) might be inclined to increase complexity in his/her writing which would backfire and cause teachers to have lower opinions of the student's intelligence. Likewise leaders facing crucial decisions might use more complex vocabulary and end up undermining others' confidence in their leadership ability. Thus it may be worthwhile to investigate ways of either preventing the tendency to use needless complexity, or look at ways that fluency biases might be overcome.

In the interim, we can conclude one thing. The pundits are likely right: write clearly and simply if you can, and you'll be more likely to be thought of as intelligent.

ACKNOWLEDGEMENTS

This material is based on work supported under a National Science Foundation Graduate Research Fellowship. The author thanks Chip Heath, Michelle Keller, Joel Allan, Busayo Ojumu, Jessica Laughlin, Norbert Schwarz, Bruce Whittlesea, Colleen Kelley, Stephen Lindsay, James Pennebaker, Benoit Monin, Herb Clark and the SLUGs, and several anonymous reviewers for advice and support.

REFERENCES

- Bem, D. J. (1995). Writing a review article for Psychological Bulletin. *Psychological Bulletin*, 118, 172–177.
- Chang, P. M. Y. (1993). An institutional analysis of the evolution of the denominational system in American Protestantism, 1790–1980. Unpublished doctoral dissertation, Stanford University.
- Descartes, R. (1990). *Meditations on first philosophy* (G. Heffernan, Trans.). London: University of Notre Dame Press. (Original work published 1641).
- Descartes, R. (1993). *Meditations on First Philosophy* (S. Tweyman, Trans.). London: Routledge. (Original work published 1641).
- Einhorn, H., & Hogarth, R. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3–19.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Grice, H. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, 56, 326–338.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases*. New York: Cambridge Press.
- Kelley, H. H. (1973). The process of causal attribution. *American Psychologist*, 28, 107–128.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, 7(1), 83–104.
- Norwick, R. J., & Epley, N. (November, 2002). *Confidence as inference from subjective experience*. Talk presented at the meeting of the Society for Judgment and Decision Making, Kansas City, MO.
- Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science*, 15(2), 100–105.

- Oppenheimer, D. M., & Frank, M. C. (submitted). A rose in any other font wouldn't smell as sweet: fluency effects in categorization.
- Oppenheimer, D. M., & Monin, B. (in prep). Factors influencing spontaneous discounting of fluency in frequency judgment.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of Personality and Social Psychology*, 77, 1296–1312.
- Pennebaker, J. W., & Lay, T. C. (2002). Language use and personality during crises: analysis of Mayor Rudolph Giuliani's press conferences. *Journal of Research in Personality*, 36, 271–282.
- Publication Manual of the American Psychological Association* (4th ed.). (1996). Washington, D.C.: American Psychological Association.
- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8, 338–342.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9, 45–48.
- Schwarz, N. (2004). Meta-cognitive experiences in consumer decision making. *Journal of Consumer Research*, 14(4), 332–348.
- Spearman, C. (1904). General intelligence, objective determined and measured. *American Journal of Psychology*, 15, 201–293.
- Steele, C. M. (1997). A threat in the air: how stereotypes shape the intellectual identities and performance of women and African-Americans. *American Psychologist*, 52, 613–629.
- Strunk, W., Jr., & White, E. B. (1979). *The elements of style* (3rd ed.). New York: Macmillan.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76, 105–110.
- Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, 5(2), 207–232.
- Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? The unexpected basis of feelings of familiarity. *Acta Psychologica*, 98, 141–166.
- Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancy-attribution hypothesis I: the heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 3–13.
- Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancy-attribution hypothesis II: expectation, uncertainty, surprise and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 14–33.
- Wilson, T. D., & Brekke, N. C. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117–142.

APPENDIX: EXAMPLES OF EXCERPTS USED AS STIMULI

Excerpts from graduate admissions essay (Experiment 1)

Original

- 1) I want to go to Graduate School so that I can learn to know literature well. I want to explore the shape and the meaning of the novel and its literary antecedents. I want to understand what the novel has meant in different literary periods, and what is likely to become. I want to explore its different forms, realism, naturalism and other modes, and the Victorian and Modernist consciousness as they are revealed.
- 2) Gold is not always a shifting, malleable metal; it is hardened by alloying with other metals, increasing its strength. I hope to go through a corresponding process at Stanford. I want to become a more solid citizen through exposure to other viewpoints and cultures, and by offering my own. I will mix with new perspectives; I will alloy with my fellow students, with my professors, and with the learning that both groups impart in order to become stronger academically, socially, and culturally.

Moderate complexity (every 3rd applicable word lengthened)

- 1) I want to go to Graduate School so that I can learn to recognize literature well. I want to explore the character and the meaning of the novel and its literary antecedents. I desire to understand what the novel has represented in different literary periods, and what is likely to become. I desire to explore its different manners, realism, naturalism and other modes, and the Victorian and Modernist consciousness as they are revealed.
- 2) Gold is not always a shifting, malleable metal; it is consolidated by alloying with other metals, increasing its strength. I hope to go through a corresponding development at Stanford. I want to become a firmer solid citizen through exposure to other perspectives and cultures, and by offering my own. I will mix with novel perspectives; I will alloy with my fellow students, with my professors, and with the knowledge that both groups impart in order to become stronger academically, communally, and culturally.

High complexity (every applicable word lengthened)

- 1) I desire to go to Graduate School so that I can learn to recognize literature satisfactorily. I want to investigate the character and the connotation of the narrative and its literary antecedents. I desire to comprehend what the narrative has represented in numerous literary periods, and what it is expected to become. I desire to investigate its numerous manners, realism, naturalism, and other approaches, and the Victorian and Modernist consciousness as they are discovered.
- 2) Gold is not constantly a changing, malleable metal; it is consolidated by alloying with additional metals, increasing its strength. I anticipate to go through a corresponding development at Stanford. I yearn to develop into a firmer substantial citizen through introduction to other perspectives and cultures, and by contributing with my own. I will combine it with novel perspectives; I will alloy with my associate scholars, with my professors, and with the knowledge that both groupings communicate in order to become stronger academically, communally and culturally.

**EXCERPT OF DIFFERENT TRANSLATIONS OF DESCARTES MEDITATION
IV (EXPERIMENT 2)**

From Tweyman's (1993) translation

'Many other matters respecting the attributes of God and my own nature or mind remain for consideration; but I shall possibly on another occasion resume the investigation of these. Now (after first noting what must be done or avoided in order to arrive at a knowledge of the truth) my principal task is to endeavor to emerge from the state of doubt into which I have these last days fallen, and to see whether nothing certain can be known regarding material things'.

From Heffernan's (1990) translation

'There remain to be investigated by me many things concerning the attributes of God, and many things concerning me myself or the nature of my mind. But I shall perhaps resume these things at another time, and now nothing seems to be more urgent (after I have noticed against what were to be cautioned and what were to be done in order to reach the truth)

than that I might try to emerge from the doubts into which I have gone in the pervious days and that I might see whether something certain concerning material things could be had'.

DISSERTATION ABSTRACT EXCERPTS (EXPERIMENT 3)

Original

This dissertation presents a historical study of the institutional development of the American religious sector. Through the lens of institutionalist perspectives developed in organizational sociology I focus on the co-evolution of the modern denominational form and the denominational system in the United States from 1790 to 1980. Through an empirical study of American Protestant denominations I build arguments which advance three theoretical issues within institutional theory.

Simplified

This thesis presents a historical study of the societal advance of the American religious sector. Through the lens of social institution views developed in organizational sociology I focus on the co-evolution of the modern denominational form and the denominational system in the Unites States from 1790 to 1980. Through an empirical study of American Protestant denominations I build arguments which advance three theoretical issues within social theory.