Surprised Elaboration: When White Men Get Longer Sentences

Lauren Eskreis-Winkler¹ and Ayelet Fishbach²

¹Kellogg School of Management, Northwestern University
²Booth School of Business, University of Chicago

In press, Journal of Personality and Social Psychology

Author Note

Lauren Eskreis-Winkler, Kellogg School of Management, Northwestern University; Ayelet Fishbach, Booth School of Business, University of Chicago.

Correspondence should be addressed to Lauren Eskreis-Winkler, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208 (email: lauren.eskreis-winkler@kellogg.northwestern.edu) and Ayelet Fishbach, Booth School of Business, University of Chicago, 5807 S. Woodlawn Avenue, Chicago, IL 60637 (email: Ayelet.Fishbach@chicagobooth.edu).

All data and materials are available on OSF: https://osf.io/h2tbe/?view_only=e28edde05cc8450698a9cdef1d6b10a58

We are grateful to the participants who made this work possible and to Carman Fowler and Eda Erensoy for their help. This work was supported by funding from the University of Chicago, Booth School of Business. L. Eskreis-Winkler extends special thanks to Ari Lustig, whose wonderfulness is so exceptional as to be surprising, and merits much elaboration.

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/pspa0000297
Abstract
We present a new consequence of stereotypes: they affect the length of communications. People say more about events that violate common stereotypes than those that confirm them, a phenomenon we dub *surprised elaboration*. Across two public datasets, government officials wrote longer reports when negative events befell white people (stereotype-inconsistent) than when the same events befell Black or Hispanic people (stereotype-consistent). Officers authored longer missing child reports of white (vs. Black or Hispanic) children (Study 1a), and medical examiners wrote longer reports of unidentified white (vs. Black or Hispanic) bodies (Study 1b). In follow-up experiments, communicators found stereotype-inconsistent events more surprising and this prompted them to elaborate (Study 2). Surprised elaboration occurred for negative events (i.e., crimes, misdemeanors), and also positive ones (i.e., weddings) (Study 3). We found that surprised elaboration has policy implications. Observers preferred to funnel government and media resources towards white victims, since their case reports were longer, even when longer reports were not more informative (Studies 4-6). Together, these studies introduce surprised elaboration, a new theoretical phenomenon with implications for public policy.

*Keywords*: surprise; elaboration; communication; stereotypes
Surprised Elaboration: When White Men Get Longer Sentences

One of the most troubling inequalities in the criminal justice system is that among Black and white offenders who commit the same crime, whites get shorter sentences (US Sentencing Commission, 2017). In the current investigation, we explore how stereotypes affect a different type of sentence: the kind that ends with a period.

We predict that people say more—that is, communicate more words—when discussing stereotype-inconsistent events (i.e., a negative event befalls a white person; a positive event befalls a Black person) than stereotype-consistent ones (i.e., a negative event befalls a Black person; a positive event befalls a white person), because they find the former more surprising. We use the term *surprised elaboration* to describe this tendency to say more about stereotype-inconsistent events. We also explore whether the number of words communicated about a topic, independent of the content itself, influences the importance readers attribute to the topic, how much they attend to it, how much they care about it, and their interest in directing resources towards it.

Quantifying Communication

The study of language typically concerns what people say. Historically, psychiatrists, psychoanalysts, and sociolinguists have argued that language provides a window into one’s psychology, if not the soul (Freud, 1901; Tausczik & Pennebaker, 2009). Indeed, peoples’ opinions, beliefs, and motives inform the topics they discuss (An et al., 2014; De Angelis et al., 2012; Wojnicki & Godes, 2008), just as their inner motivations shape the arguments they make and the conclusions they draw (Kunda, 1990). In recent years, developments in machine learning have led to an explosion of language analysis tools that quantify the topics, tone, and sentiment of language at scale (Bhatia, Mellers, & Walasek, 2019; Schwartz & Ungar, 2015).

The content of communication, such as the ideas, beliefs, and thoughts it contains, is no doubt its central feature. Yet sometimes, ancillary features of language are important in their own right. In one classic study, individuals who asked to cut in line for a copy machine were more successful if they accompanied the request with a reason (“May I use the xerox machine, because I have to make copies?”), even though the reason was devoid of information (Langer et al., 1978). So too, conversational hedges (i.e., “somewhat” or “might”) reflect key features of the communicator’s personality and can affect how recipients respond to messages (Yeomans et al., 2020), though they contain no real content.
Contributing to this body of work, we investigate another ancillary feature of language: how much people say. We examine whether the number of words communicated indexes the speaker’s surprise.

When information matches peoples’ expectations, people do not exhibit strong emotional or cognitive reactions—in fact, it sometimes goes unnoticed. However, when information is unexpected, it engenders surprise (Meyer et al., 1991; Reisenzein, Horstmann & Schützwohl, 2019). Surprise focuses attention since people are motivated to make sense of things that violate their expectations (Foster & Keane, 2015; Hutter & Crisp, 2005; Hutter et al., 2009; Itti & Baldi, 2009; Kahneman & Miller, 1986). Surprising events lead people to wonder “why” an event occurred (Wong & Weiner, 1981) which prompts deeper information processing (Hastie, 1984; Pyszczynski & Greenberg, 1981).

How people process surprising information has a direct effect on how people communicate about it. Out of a desire to make sense of surprising events, people are more likely to try to explain them. For example, journalists write more in-depth articles when reporting on unexpected sporting outcomes (i.e., the underdog wins; Lau & Russell, 1980). In lab experiments, surprise has been shown to prompt explanation. In one set of studies, participants were more likely to complete a sentence stem that violated expectations (e.g., “The youngest kid in the class received an A on the test. . .”) with an explanation (e.g., “. . .because he studied hard”) as compared to when the sentence stem confirmed expectations (Hastie, 1984; Sekaquaptewa, Espinoza, Thompson, et al., 2003; Von Hippel et al., 1997).

In the domain of language and communication, relatively little attention has been given to another way in which violated expectations might affect the communicator: by prompting engagement. Indeed, surprise is known to affect engagement. The novelty-complexity of an event—that is, how unexpected, surprising, complex, and difficult to understand it is—piques interest (Berlyne, 1960; Silvia, 2006). Given the choice between looking at a red strawberry or a blue one, people are more surprised by the blue one, more interested by it, and therefore engage with it more. Just like infants who stare longer at unusual stimuli (Fantz, 1964), students pay more attention to unexpected materials (Hidi, 1995; Sadoski, 2001), and observers gaze at strange pictures more (Silvia, 2005). Evolutionary scientists have proposed that interest has a functional value: it motivates people to engage with the environment (Izard & Ackerman, 2000).
People use this surprise-engagement association strategically. For example, students devise strategies to boost their interest to fend off disengagement (Sansone & Harazkiewicz, 1996). Storytellers have long intuited that surprising, unexpected events capture attention and keep it, which is why the most engaging stories contain an element of surprise (Heath & Heath, 2007). Given the psychological pull to resolve curiosity and answer open questions (Kruglanski & Webster, 1996; Shen et al., 2015), surprising events draw engagement. Modern TV advertisers use this fact to their advantage, for example by creating TV ads that only reveal the name of the product at the end of the commercial to hold the viewer’s attention (Loewenstein, 1994).

Surprise may similarly motivate engagement among communicators. When communicators encounter surprising stimuli, the surprise may motivate them to pay attention longer, to think more about the stimulus, and therefore, to say more about it. While the desire to make sense of a surprising event and explain it could, in and of itself, lead to greater engagement, we propose that surprise motivates engagement even in the absence of explicit sense-making or explanatory search. Thus, independent of the effect surprise has on reasoning processes—i.e., by triggering sense-making or attributional search—a communicator may engage more with surprising events, and as a result, say more about them.

**The Surprised Elaboration Effect**

We expect people to say more about things that are unexpected. Accordingly, our first hypothesis is that communicators will say more about events that violate (vs. confirm) expectations, a phenomenon we refer to as surprised elaboration. We operationalize elaboration as word count, predicting that when an event violates expectations, communicators will write more words about it.

We theorize that, in general, violated expectations prompt elaboration—for example, a violated expectation about the weather, the economy, or how a supermarket cashier ought to behave. In this research, we explore surprised elaboration in the wake of one specific type of violated expectations: common stereotypes.

Stereotypes are at the core at human perception. Racial stereotypes in particular guide group perception. Disadvantaged minority groups are, broadly, negatively stereotyped. For example, Black and Hispanic people in the US are commonly assumed to be poor, unhealthy, uneducated, and dangerous (Axt & Trawalter, 2017; Eberhardt, 2020; Fiske et al., 2002; Goff et al., 2008; Zou & Cheryan, 2017). People hold this overgeneralized, a priori belief about
disadvantaged minorities partially due to illusory correlations: the cognitive tendency to associate things they believe to be rare (i.e., violence) with minority populations (i.e., Blacks, Hispanics) (Hamilton & Rose, 1980; McConnell et al., 1994; Mullen & Johnson, 1990).

We test whether stereotype-inconsistent (versus stereotype-consistent) events prompt surprised elaboration. For example, we examine whether officers write longer report for missing white children than missing Black children because officers negative events are less likely to befall white (vs. Black) children.

Moving to the underlying, psychological process, our second hypothesis is that violated stereotypes prompt elaboration because they are surprising. Since Black and Hispanic people are often stereotyped in negative ways, we expect negative events that befall white (vs. Black or Hispanic) individuals to seem more surprising and that positive events that befall white (vs. Black or Hispanic) people will seem less surprising. That is, we predict communicators will say more for a white (vs. Black) person who is found dead, but more for a Black (vs. white) person who wins an award.

Surprise can originate from many sources. One is the rarity of an event. Rare things are more surprising than common ones. Alternatively, surprise can occur when general expectations—such as general stereotypes—are violated, even if the event was not statistically infrequent. For example, a four-leaf clover is surprising because most clovers have just three leaves. In contrast, a white child who goes missing is more surprising than a Black child who goes missing because this child violates the stereotype that negative events are more likely to befall minorities. Yet it is not statistically true that most kids who go missing are nonwhite.

We theorize that violated stereotypes generate surprise when people do not have access to information on the commonness or rarity of an event. For rare events (e.g., a child goes missing), the police officer who reports the case has limited personal experience reporting on missing children but can nonetheless be guided by a general racial stereotype that bad things are less likely to happen to whites (vs. Blacks). As a result, we expect the officer to elaborate more on the white missing child than the missing Black one.

We also explore what exactly people elaborate on when they say more. As noted, surprise might lead people to generate explanations (Hastie, 1984; Sekaquaptewa, Espinoza, Thompson, et al., 2003; Von Hippel et al., 1997) and thereby author longer texts. Yet, we expect that surprised elaboration does not depend on explanations. Even when communicators do not engage
in sense-making, we expect them to simply engage longer with surprising stimuli and, as a result, say more about them. To test this, we examined surprised elaboration in texts where explanations tend to be absent (i.e., missing child police reports, unidentified body reports). In these contexts, we tested whether communicators who are surprised write more detail (e.g., describe more clothing items a missing child was last seen wearing) and are more specific in their language (e.g., describe the brand name for the child’s jeans), but not engaged in sense-making.

What are the implications of surprised elaboration? Seemingly innocuous behaviors, such as the rate at which professors reply to emails (Milkman, Akinola, & Chugh, 2015), the appellations people use to refer to others (Atir & Ferguson, 2018), and the designs of virtual classrooms (Cheryan, Meltzoff, & Kim, 2011) can perpetuate social inequalities. Similarly, surprised elaboration might influence how people respond to communication.

On the one hand, observers might react negatively to longer text. Longer text could be taken as a sign of disorganization or lack of professionalism on the part of the communicator. Longer text might also be viewed negatively if the reader perceives themselves being coerced by the author to reach a certain conclusion.

In contrast, longer texts could signal greater importance. Our third and final hypothesis is that observers will infer that reports with longer text are more important, and therefore they be more interested in investing resources in these cases. Combined with the first hypothesis, this suggests that people will prefer to invest resources in the cases of white victims (who have long case reports) versus Black victims (who have shorter case reports). Notably, this third hypothesis does not suggest racial bias. Even if an allocator does not prefer to invest in one social group over another, or even prefers to invest more in minorities, the allocator who heuristically invests in cases with longer text may end up choosing to invest more in white (vs. Black and Hispanic) reports if the former cases tend to be longer.

Beyond surprise, we explore an alternative process that could account for surprised elaboration: in-group favoritism. People tend to favor similar others, caring more about members of their in-group than out-group (Tajfel, Billig, Bundy, & Flament, 1971). In-group favoritism could lead predominantly white communicators to say more about white (vs. Black or Hispanic) victims in need of help if they feel more familiar with in-group members, have more compassion for in-group members, and are more motivated to service them well.
To address this alternative, we test for surprised elaboration in the wake of both negative and positive events. Following positive events, a surprise mechanism and an in-group favoritism mechanism make opposite predictions. Whereas an in-group favoritism mechanism predicts that white communicators will say more about white (vs. Black or Hispanic) targets, a surprise mechanism predicts that communicators will find positive events that befall white (vs. Black or Hispanic) individuals less surprising, and as a result, will say less about them. For example, in the context of another stereotype—that same-sex relationships are weak and therefore do not lead to marriage (Peplau & Fingerhut, 2007; Testa et al., 1987)—a surprise mechanism predicts that reporters will find same-sex weddings more surprising and say more about them in news articles, whereas an ingroup favoritism predicts that reporters (the majority of whom are likely heterosexuals, based on population base rates) will write shorter articles, elaborating less.

The Present Research

We test for evidence of surprised elaboration across publicly-available datasets. We began by examining whether officers write longer missing child reports for white (vs. Black and Hispanic) children, and whether medical examiners write longer descriptions of unidentified white (vs. Black and Hispanic) bodies (Study 1). In testing for surprised elaboration in these datasets, we searched for evidence of this phenomenon in the real world: that is, we examined whether real-world stereotype-inconsistent (vs. stereotype-consistent) events prompt greater elaboration.

Complementing these archival analyses, in follow-up experiments we tested for a causal path from stereotype-inconsistent events to greater elaboration, and examined whether surprise mediated this effect (Study 2). Next, in both a lab experiment and a field study, we explored whether the association between surprise and elaboration generalized to positive events—like winning an award or having a wedding (Study 3).

Finally, we examined public policy implications: whether people prefer to invest resources in cases accompanied by longer reports—even when longer reports contain no additional information (Studies 4-6). If so, this suggests that observers may inadvertently funnel resources towards white (vs. Black and Hispanic) victims since the former tend to have longer case reports, even if the observers’ explicit preference is to be egalitarian. Moreover, this would suggest that text length is an important feature of language and communications in its own right. A summary of the studies and their key results are listed in Table 1.
Across experiments, to maximize power, we targeted a minimum sample of 50 participants per cell. Based on preliminary pilots, we anticipated that we would find medium to large effects. Power analyses conducted in G*Power based on respective sample sizes and target alpha level ($\alpha = .05$) revealed that power was sufficient across all studies (i.e., $\geq .80$) to detect a medium to large effect (e.g., $d = .60, \eta^2_p = .08$). The studies in this paper incorporate data from participants in the field and online participants recruited via Amazon Mechanical Turk (MTurk). For data and materials on all experiments, see OSF (https://osf.io/h2tbe/?view_only=e28ede05cc8450698a9c0f1d6b10a58).

**Table 1**

*Summary of Results*

<table>
<thead>
<tr>
<th>Study</th>
<th>Main Findings</th>
</tr>
</thead>
</table>
| 1     | 1a. Officers wrote longer reports for white (vs. minority) children who went missing.  
       | 1b. Medical examiners wrote longer reports for white (vs. minority) unidentified bodies. |
| 2     | Participants who learned about a crime committed by a white (vs. Black) individual were more surprised. Surprise, in turn, predicted writing more about the incident. Follow-Up: The effect of surprise on elaboration replicated in a sample of all-Black participants. |
| 3     | Generalizing to positive events:  
       | 3a. Participants wrote more about a positive incident involving a Black (vs. white) individual—and a negative incident involving a white (vs. Black) individual.  
       | 3b. *New York Times* articles about same-sex (vs. heterosexual) couples were longer. |
| 4     | Reviewing unidentified body reports, participants chose to invest more resources in lengthier texts. |
| 5     | Participants chose to invest more resources in lengthier missing child and unidentified body reports, even when the extra information was irrelevant. |
| 6     | Participants chose to invest more resources in lengthier missing child reports, even when the extra information was illegible. |

**Study 1: Surprised Elaboration in the Field**

In Study 1, we examined whether government officials wrote longer reports when white (vs. Black or Hispanic) individuals experienced negative outcomes. We hypothesized that officials would engage more with stereotype-inconsistent outcomes, and as a result, write longer reports of these incidents. We analyzed elaboration patterns in two publicly-available datasets. Study 1a explored whether officers wrote longer missing child reports when white (vs. Black or Hispanic) went missing. Study 1b examined whether medical examiners wrote longer reports for
white (vs. Black or Hispanic) unidentified bodies. Across both studies, we conducted textual analysis to better understand the content contained in longer text.

Prior to testing the main hypothesis in Study 1, we conducted a pretest to ensure that people do, in fact, find stereotype-inconsistent events surprising. In this pretest, we found that people are more surprised when various negative events—a child goes missing, a body is unidentified, a drug offense—happen to white (vs. Black) people, and when positive events—i.e., winning an award—happen to Black (vs. white) people (see Supplement for methods and results).

**Study 1a: White Missing Children Get Longer Reports**

We predicted that officers would author longer reports for missing white (vs. Black or Hispanic) children. To test this, we analyzed the narrative descriptions of real-world missing children’s posters. As a complementary analysis, we coded the text of the reports to better understand what additional content is contained in lengthier reports.

**Method**

**Materials and Procedure.** Posters of children who are actively missing were retrieved from the National Center for Missing and Exploited Children’s (NCMEC) online database, a non-profit organization established by an act of Congress. Federal law requires that States report missing children to law enforcement agencies and the NCMEC. For a child to be eligible for a missing child poster, the child needs to be registered in the FBI’s database as a missing child. The NCMEC collaborates with parents, guardians, and law enforcement agencies to create the narrative descriptions on the missing child posters. Posters for children who have been found are removed from this public database. Each actively missing individual has a poster on the website that contains both a physical description (i.e., gender, race, hair color, eye color) and a narrative description of the missing child.

Our sample consisted of all publicly-available missing child posters ($N = 1,051$) for white (32.1%), Black (21.4%) and Hispanic (43.7%) children who had gone missing in the four largest US states: California (38.8%), Texas (25.8%), Florida (23.2%) and New York (12.1%). We included posters of all children who had gone missing over the past 30 years (62.3% female). The percentage of missing children in our dataset who were white (32.1%) approximates the percentage of white kids in the four states we studied (ranging from 30-50%; Kids Count Data Center). Thus, it is not the case that Black and Hispanic children are much more or much less
likely than white children to go missing across the four states. Nevertheless, if people stereotype, believing Black and Hispanic minorities are more likely to experience negative outcomes, then they should be more surprised when white children go missing.

**Measures.** The downloaded dataset contained the following variables, which we analyzed.

**Elaboration.** Each missing child poster has a narrative description of the missing child. We calculated a word count for each of these narrative descriptions, which served as the primary dependent variable.

**Demographics.** Each missing child report has demographic information on the missing child. We used the following variables as predictors and/or covariates in our analyses: age when the child went missing, race of the missing child, gender of the missing child, child’s state of origin, and date when the child went missing.

**Coded variables.** Two coders coded the text of the reports for three variables. First, number of explanations \( r = .85 \). Anything in the report that explained why the child was missing was coded as an explanation (for example, “She was abducted by her grandparents…”). Second, types of details \( r = .94 \). Coders coded the number of different types of details (i.e., date the child was last seen, what the child was last seen wearing) the report contained. Finally, coders coded for specificity \( r = .81 \), using a scale from 1 = *not at all specific*, to 5 = *incredibly specific*. “Missing kid wore jeans” for example, was not as specific as “Missing kid wore blue Levi jeans with wide leg and was last seen at home at 1:21 PM on 3/2/2018.”

**Results**

On average, missing children were around 14 years old when they went missing \( M_{\text{age}} = 14.47, SD = 4.70 \). The county from which the child went missing was reported in 82% of cases.

An ANOVA of race (White vs. Black vs. Hispanic) on word count revealed significant differences between report lengths of white children \( M = 29.93, SD = 20.79 \), Black children \( M = 20.84, SD = 14.38 \) and Hispanic children \( M = 23.10, SD = 17.64 \), \( F(2, 1,048) = 21.85, p < .001, \omega^2 = .040 \). Consistent with our hypothesis, white children received longer reports than Black and Hispanic children, \( F(1, 1,049) = 41.06, p < .001, \omega^2 = .038 \). (For a summary of the surprised elaboration effect across Studies 1-3, see Table 2.) This effect remained significant when controlling for four fixed effect variables available in the dataset (the child’s state of origin, gender, age, and the date the child went missing) and one random effect (the county from
which the child went missing), $β = -3.51, SE = 1.01$, 95% CI: $-5.49 - -1.52$, $t(941.30) = -3.47, p = .001$. For intercorrelations among these variables, as well as full details on all models, see Tables S1-S2 in the Supplement.¹

Table 2

Summary of the Surprised Elaboration effect across Studies 1-3

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Problem domain(s)</th>
<th>Stereotype-consistent event (word count)</th>
<th>Stereotype-inconsistent event (word count)</th>
<th>Surprised elaboration effect</th>
<th>Effect Size</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1a</td>
<td>1,051</td>
<td>Missing Children</td>
<td>22.30 words</td>
<td>29.93 words</td>
<td>$F(1, 1,049) = 41.06, p &lt; .001$</td>
<td>$d = 0.42$</td>
<td>[0.29, 0.55]</td>
</tr>
<tr>
<td>Study 1b</td>
<td>1,194</td>
<td>Unidentified Bodies</td>
<td>49.84 words</td>
<td>58.80 words</td>
<td>$F(1, 1,192) = 7.46, p = .006$</td>
<td>$d = 0.16$</td>
<td>[0.04, 0.27]</td>
</tr>
<tr>
<td>Study 2</td>
<td>206</td>
<td>Drug Offense</td>
<td>8.35 words</td>
<td>12.70 words</td>
<td>$t(204) = 2.31, p = .022$</td>
<td>$d = 0.32$</td>
<td>[0.05, 0.60]</td>
</tr>
<tr>
<td>Follow-Up</td>
<td>200</td>
<td>Drug Offense</td>
<td>4.22 words</td>
<td>17.62 words</td>
<td>$t(198) = 4.32, p &lt; .001$</td>
<td>$d = 0.61$</td>
<td>[0.33, 0.90]</td>
</tr>
<tr>
<td>Study 3a</td>
<td>402</td>
<td>Teacher Award</td>
<td>24.77 words</td>
<td>33.89 words</td>
<td>$F(1, 398) = 8.76, p = .003$</td>
<td>$d = 0.30$</td>
<td>[0.10, 0.50]</td>
</tr>
<tr>
<td>Pos. event</td>
<td>402</td>
<td>Harassment</td>
<td>39.47 words</td>
<td>47.38 words</td>
<td>$F(1, 398) = 6.64, p = .010$</td>
<td>$d = 0.25$</td>
<td>[0.06, 0.45]</td>
</tr>
<tr>
<td>Neg. event</td>
<td>402</td>
<td>NYT Wedding Articles</td>
<td>517.69 words</td>
<td>731.29 words</td>
<td>$F(1, 481) = 6.48, p = .011$</td>
<td>$d = 0.39$</td>
<td>[0.09, 0.69]</td>
</tr>
</tbody>
</table>

¹ There was no effect for gender on elaboration. Thus, while the majority of missing kids were girls (62.3% vs. 37.7% boys), the gender of the child did not predict length of the report, $F(1, 1,049) = .017, p = .679$, likely because a violation of pure statistical probability—absent the violation of expectations—does not motivate elaboration. In contrast, age in this sample motivated elaboration. The age of a missing child was negatively correlated with the length of the narrative description in his/her report, $r = -.40, p < .001$, suggesting that, indeed, when younger children go missing this defies expectations and motivates elaboration. In a follow-up study, participants thought it was more surprising to have a missing white (vs. Black) child or a missing young (vs. old) child than a missing girl (vs. boy). In other words, people have expectations that white (vs. Black) and young (vs. old) children are less likely to go missing, but they do not have an expectation about whether girls versus boys are more likely to go missing. For details, see the Supplement.
Analyses of text quality. Across the three coded variables, coders’ ratings were correlated at or above $r = .81$. We ran an ANOVA of race (White vs. Black vs. Hispanic) on each of the three coded variables: explanations, detail, and specificity. An ANOVA of race (White vs. Black vs. Hispanic) on explanations revealed significant differences between the number of explanations in the reports of white children ($M = 0.08, SD = 0.27$), Black children ($M = 0.04, SD = 0.19$) and Hispanic children ($M = 0.05, SD = 0.21$), $F(2, 1,048) = 3.07, p = .047, \omega^2 = .006$. An ANOVA of race (White vs. Black vs. Hispanic) on details also revealed significant differences between how many types of detail were contained in the reports of white children ($M = 5.34, SD = 3.74$), Black children ($M = 3.64, SD = 2.56$) and Hispanic children ($M = 4.06, SD = 3.17$), $F(2, 1,048) = 23.69, p < .001, \omega^2 = .043$. Finally, an ANOVA of race (White vs. Black vs. Hispanic) on specificity similarly revealed significant differences between the specificity of the reports of white children ($M = 3.05, SD = 1.25$), Black children ($M = 2.44, SD = 1.00$) and Hispanic children ($M = 2.58, SD = 1.10$), $F(2, 1,048) = 25.91, p < .001, \omega^2 = .047$.

Next, we examined whether explanation, detail or specificity mediated the effect of race on elaboration. All three coded variables independently mediated the effect of race on word count, $\beta_{\text{indirect\_Explanations}} = 1.06, SE = 0.44, 95\% \text{ CI} [0.15, 1.95]; \beta_{\text{indirect\_Detail}} = 7.66, SE = 1.15, 95\% \text{ CI} [5.30, 10.13]; \beta_{\text{indirect\_Specificity}} = 7.84, SE = 1.12, 95\% \text{ CI} [5.56, 10.22]$ (all models based on 10,000 bootstrap samples). However, only specificity remained a significant mediator when all coded variables were added to the model, $\beta_{\text{indirect\_Specificity}} = 0.25, SE = 0.12, 95\% \text{ CI} [0.02, 0.48]$. We conclude that the effect of race on longer text is driven by communicators choosing to describe the event using greater specificity. See Supplement for full details on these and other models.

**Study 1b: White Unidentified Bodies Get Longer Reports**

In Study 1a, the information in missing child reports was created in collaboration with parents and other family members and friends of the missing individual. In Study 1b, we analyzed the length of the narrative descriptions authored by different government officials: medical examiners. Of necessity, reports for unidentified bodies are typically written without input from anyone connected to the body, making them a more direct reflection of the author’s expectations. We predicted that medical examiners would write longer reports for white (vs. Black or Hispanic) unidentified bodies. As in Study 1a, we coded the content of the report to see what makes reports longer.
**Method**

**Materials and Procedure.** We downloaded medical examiner reports for the 2,000 most recent unidentified bodies listed in the National Missing and Unidentified Persons System (NamUs) database (2020). We went back six years in the database in order to gather 2,000 reports. NamUs, a national clearinghouse administered by the National Institute of Justice, stores medical examiner records on unidentified persons. Medical examiner reports contain basic facts about the unidentified body—for example, the circumstances in which the body was discovered and what was found on the body. The purpose of these reports is to aid in the identification of unidentified bodies.

Our sample consisted of narrative reports for the 1,194 bodies with an identifiable ethnicity that were either white (48.1%), Black (14.0%) or Hispanic (37.9%). We note that “Hispanic” is a cultural category, not a racial designation. Nevertheless, medical examiners were able to identify some bodies as Hispanic because in some cases, a third party knew something about the unidentified body that suggested the target was Hispanic. For example, maybe a neighbor knew part of the person's name (e.g., “*Hispanic male found unresponsive...[p]robable name of Carlos*”), or where the person was from (e.g., “…*decedent was found unresponsive...[n]eighbors indicate the decedent is....from Mexico*”), which led the medical examiner to identify the body as Hispanic.

**Measures.** The downloaded dataset contained the following variables, which we analyzed.

**Elaboration.** Each unidentified body report had a narrative description of the unidentified body. We calculated a word count for each of these narrative descriptions, which served as the primary dependent variable.

**Demographics.** Each unidentified body report contained demographic information on the body. We used the following variables as predictors and/or covariates in our analyses: race of the unidentified body, gender of the unidentified body, date when the body was found, and zip code in which the body was found—although this zip code variable was only present for 38% of the bodies in the dataset.

**Coded variables.** Two coders coded each report for the same three variables that were coded in Study 1a: number of explanations \( r = .63 \), number of different types of detail \( r = .80 \), and specificity \( r = .88 \). See Study 1a for details.
Results

An ANOVA of race (white vs. Black vs. Hispanic) on word count revealed significant differences between report lengths of white bodies ($M = 58.80, SD = 64.58$), Black bodies ($M = 49.18, SD = 39.19$), and Hispanic bodies ($M = 50.09, SD = 51.08$), $F(2, 1,191) = 3.74, p = .024, \omega^2 = .006$. Consistent with the hypothesis, white bodies received longer reports than Black and Hispanic bodies, $F(1, 1,192) = 7.46, p = .006, \omega^2 = .01$. This effect remained significant when controlling for fixed effect variables in the dataset (gender, the date the body was found), $F(1, 1,170) = 7.01, p = .008, \omega^2 = .01$. For intercorrelations among these variables and full details on all models, see Tables S3-S4 in the Supplement.2

One variable that might confound the main effect is income: it is possible that white unidentified bodies were found in better-resourced neighborhoods, where better-funded police departments had more time to spend on each case and therefore wrote longer reports. To address this possibility, we tested whether the median income level of the zip code where the body was found predicted report length. Recall that only a small percent of the bodies (38%) in the dataset reported the zip code in which the body was found. Matching these zip codes to records from the U.S. Census (2018), we calculated the median income for each zip code. In the subsample of bodies with zip codes, the median income was $47,361.09 (SD = 23, 796.45). Median income level was not related to the word count of the report, $r = .02, p = .629$.

Analyses of text quality. An ANOVA of race (white vs. Black vs. Hispanic) on explanations revealed no significant differences between the number of explanations in the reports of white children ($M = 0.08, SD = 0.25$), Black children ($M = 0.06, SD = 0.23$) and Hispanic children ($M = 0.06, SD = 0.21$), $F(2, 1,187) = 0.80, p = .451, \omega^2 = .001$. In contrast, an ANOVA of race (white vs. Black vs. Hispanic) on details revealed significant differences between how many types of detail were contained in the reports of white children ($M = 4.66, SD = 3.44$), Black children ($M = 4.46, SD = 2.93$) and Hispanic children ($M = 3.91, SD = 2.33$), $F(2, 1,1187) = 8.09, p < .001, \omega^2 = .013$. Finally, an ANOVA of race (white vs. Black vs. Hispanic)

---

2 In a complementary gender analysis, we found that the majority of unidentified bodies were male (82.5%) as opposed to female (15.8%). Nevertheless, we neither predicted nor found an effect for gender on length of the report, $F(1, 1,172) = 0.00, p = .990$, suggesting that a common stereotype needs to be violated—not just pure statistical probability—to motivate elaboration. Indeed, in a follow-up study, participants thought it was more surprising to come across an unidentified white (vs. Black) body than an unidentified female (vs. male) body. In other words, people have expectations that white (vs. Black) bodies are less likely to be found unidentified, they do not have expectations about whether female versus male bodies are more likely to be found unidentified. For details, see the Supplement.
on specificity revealed significant differences between the specificity of the reports of white
children ($M = 3.34, SD = 1.16$), Black children ($M = 3.33, SD = 1.17$) and Hispanic children ($M$
$= 3.04, SD = 1.10$), $F(2, 1,187) = 10.66, p < .001, \omega^2 = .018$.

Next, we examined whether detail or specificity mediated the effect of race on
elaboration. Both detail and specificity independently mediated the effect of race on word count,
$\beta_{\text{indirect Detail}} = 6.33, SE = 1.85, 95\% \text{ CI [2.53, 11.43]}; \beta_{\text{indirect Specificity}} = 5.04, SE = 1.46, 95\% \text{ CI [2.11, 8.36]}$ (both based on 10,000 bootstrap samples). However, neither variable remained a
significant mediator when the other was added to the model, $\beta_{\text{indirect Detail}} = 0.99, SE = 1.19, 95\% \text{ CI [-1.06, 3.89]}; \beta_{\text{indirect Specificity}} = -3.09, SE = 0.31, 95\% \text{ CI [-1.42, 0.32]}$ (both based on 10,000
bootstrap samples). We conclude that the effect of longer text is mostly driven by
communicators choosing to describe the event using greater specificity and detail, though neither
of these variables independently accounted for the overall effect. See Supplement for full details
on these and other models.

Discussion

Study 1 brought empirical evidence to support the theory of surprised elaboration. When
common stereotypes were violated, communicators elaborated. Officers and medical examiners
wrote longer reports when white (vs. Black or Hispanic) children went missing, as well as when
white (vs. Black or Hispanic) bodies were found unidentified.

Content analysis revealed that reports of white individuals were consistently more
specific and detailed than reports of Black or Hispanic individuals, but did not reliably contain
more explanations. What made these reports lengthier was therefore the amount of specificity
and detail provided.\(^3\)

\(^3\) We also conducted text analysis in a second way. We analyzed all missing child reports and unidentified body
reports from Studies 1a-1b using the Linguistic Inquiry and Word Count (LIWC) 2015 dictionary (Pennebaker,
Booth, Boyd, & Francis, 2015). LIWC 2015 created 82 language variables for each piece of text. These 82 language
variables include summary variables (i.e., “emotional tone”), linguistic dimensions (i.e., “pronouns”), psychological
processes (i.e., “position emotion”), and other grammar (i.e., “auxiliary verbs”). We tested whether any language
features generated by this program differed by race (white vs. Black or Hispanic). We did find between-condition
differences on 10 LIWC-generated language variables (i.e., percent of words that are pronouns, percent of words
that are verbs) across Studies 1a and 1b. We combined both studies and tested whether condition continued to
predict word count (the main effect) when controlling for all 10 of these variables (study was entered in the
regression as a fixed effect). The main effect remained significant which suggests that stereotype-inconsistent (vs.
stereotype-consistent events receive longer reports, and this effect cannot be explained by standard features of
language captured in LIWC. (For further details on these analyses, see the supplementary materials.)
While the archival analyses in Study 1 are consistent with a theory of surprised elaboration, field data do not provide causal evidence. In Study 2, we tested for surprised elaboration using an experimental design.

**Study 2: A Surprising Mechanism**

According to our theory, people elaborate more on stereotype-inconsistent (vs. stereotype-consistent) events because they are surprising. In Study 2 we gathered experimental evidence for this process model. We manipulated the stereotype-consistency of a negative event by telling participants that either a white or Black individual had engaged in criminal conduct. Following this, we measured how surprising people found the event, how much people chose to write about the event, and whether surprise mediated the effect of race on elaboration.

In Study 2, participants were recruited from MTurk—where samples tend to be 70-80% white (Moss & Litman, 2020). The officers from Study 1 were also likely disproportionately white, given that nationally, 67% of law enforcement officers are white (Police officers, n.d.). In a follow-up to Study 2, we tested whether surprised elaboration also occurred among Black communicators. We predicted that surprised elaboration occurs regardless of the racial match between communicator and target. This follow-up also built on Study 2 by testing for a main effect without measuring self-reported surprise. We did this to ensure that having participants report on their surprise (something they did in Study 2 but not the follow-up) is not critical for surfacing the surprised elaboration effect (i.e., via demand effects).

**Method**

**Participants**

We recruited participants on MTurk. Participants of any nationality were allowed to participate so long as their approval rating was at or above 50%. Prior to randomization to ensure that participants were motivated and paying attention, participants in this and all subsequent online studies reported in the paper also answered an open-text question: “Tell us about your favorite book and why it is your favorite book.” Using this question, unmotivated participants could drop the survey before starting the experiment. We opened the survey to 200 participants in return for $0.40; MTurk returned 206 respondents (45.5% female; $M_{\text{age}} = 38.54, SD_{\text{age}} = 12.09$).
**Procedure**

We ran a 2-condition (Race: Black vs. white) between-subjects design. Participants were introduced to a professor who had engaged in criminal conduct: “Yesterday, the community college professor pictured above was accused of coming to class while on drugs.” Whereas half of participants saw a photo of a black man, the other half of the participants saw a photo of a white man. Photos were taken from a series of publicly-available online mugshots (see materials on OSF). Participants reported how surprised people would be by this news story, from 1 = *Not at all surprised* to 5 = *Incredibly surprised*. Next, participants imagined themselves as reporters for the local newspaper, and were told to write up a short article about the incident, if they found it newsworthy. The dependent variable was word count.

**Results and Discussion**

On average, participants wrote a report that was 10.46 words ($SD = 13.63$). Recall that participants were able to write as much or as little as they wanted. Whereas 59.2% of participants chose to write something, 40.8% opted to write nothing at all. These participants got a value of 0 for number of words written.$^4$

In support of our main hypothesis, participants wrote more words about the white professor ($M = 12.70$ words, $SD = 15.52$) than the Black one ($M = 8.35$ words, $SD = 11.24$), $t(204) = 2.31$, $p = .022$, $d = .32$, $95\% CI [0.05, 0.60]$. Participants were also more surprised by the white professor’s drug offense ($M = 3.40$, $SD = 1.14$) than the offense of the Black professor ($M = 2.92$, $SD = 1.15$), $t(204) = 2.98$, $p = .003$, $d = .42$, $95\% CI [0.14, 0.70]$. Furthermore, surprise predicted words written, $B = 4.12$, $SE = 0.77$, $t(204) = 5.38$, $p < .001$. When included as simultaneous predictors in a regression, race no longer predicted word count, $B = -2.50$, $SE = 1.82$, $t(203) = -1.37$, $p = .171$, but surprise did, $B = 3.90$, $SE = .78$, $t(203) = 5.00$, $p < .001$. Next,$^4$

$^4$ Given that many participants in this study chose to write nothing, the dependent variable—word count—was not normally distributed. To account for this abnormality we first re-analyzed the data using a non-parametric Mann Whitney model; doing so did not change the significance of the main effect, $Z = -2.16$, $p = .031$. Next, we created a binary variable to indicate whether participants wrote nothing (= 0) or something (= 1). Whereas 53.77% of participants (57/106) wrote something (vs. nothing) in the Black condition, 65.00% (65/100) wrote something (vs. nothing) in the White condition, a difference that was directionally consistent with the main result but not significant: $\chi^2(1) = 2.69$, $p = .101$. Third, we analyzed the number of words written by participants in each condition, only including the $n = 122$ participants who actually wrote something in our analysis. In this subsample, participants trended towards writing more words in the white condition ($M = 19.54$ words, $SD = 15.40$) as compared to the Black condition ($M = 15.53$ words, $SD = 11.11$) but this difference did not reach significance, $t(120) = 1.63$, $p = .106$. 


we tested for mediation. Supporting our hypothesis, surprise mediated the effect of race on elaboration, $\beta_{\text{indirect}} = -1.85$, SE = .74, 95% CI [-3.40, -0.59]; based on 10,000 bootstrap samples.

**Follow-Up.** We recruited 200 Black participants from Prolific (according to demographic data on the Prolific panel; 60.0% female; $M_{\text{age}} = 29.57$, $SD_{\text{age}} = 8.93$). This study was pre-registered on AsPredicted (https://aspredicted.org/blind.php?x=uw5t86) and followed the same procedure as Study 2, except we did not measure surprise.

On average, participants wrote up a report that was 10.79 words ($SD = 22.91$). Recall that participants were able to write as much or as little as they wanted. Whereas 29.0% of participants chose to write something, the majority—71.0%—wrote nothing at all. These participants got a value of 0 for number of words written.\(^5\)

In support of the hypothesis, participants wrote more about the white professor ($M = 17.62$ words, $SD = 28.24$) than the Black professor ($M = 4.22$ words, $SD = 13.38$), $t(198) = 4.32$, $p < .001$, $d = .61$, 95% CI [.33, .89].

Moving beyond field data, in Study 2 we gathered direct evidence for the proposed process. Self-reported surprise accounted for the fact that participants wrote longer reports on stereotype-inconsistent events than stereotype-consistent ones. Moreover, the surprised elaboration effect generalized across participants, regardless of race.

**Study 3: Generalizing to Positive Events**

We next tested whether surprised elaboration would occur for positive events. If people elaborate more when stereotypes are violated, this ought to occur for an unexpected positive event just as it occurs for an unexpected negative event.

In Study 3a, we randomized participants to write about a white or a Black person experiencing a negative or a positive event. We hypothesized that people would say more about the white (vs. Black) person experiencing a negative event, yet more about the Black (vs. white)  

---

\(^5\) To account for the large number of participants who wrote nothing, we first re-analyzed the data using a non-parametric Mann Whitney model; doing so did not change the significance of the main effect, $Z = -4.90$, $p < .001$. Second, we created a binary variable to indicate whether participants wrote nothing (= 0) or something (= 1). Whereas 13.73% of participants (14/102) wrote something (vs. nothing) in the Black condition, 44.90% (44/98) chose to write something (vs. nothing) in the White condition, $\chi^2(1) = 23.59$, $p < .001$. Third, we analyzed the number of words written by participants in each condition, only including the $n = 58$ participants who actually wrote something in the analysis. Note that in the Black condition, there were a very small number of participants—only 14—included in this analysis. In this subsample, participants wrote the same number of words in the White condition ($M = 39.25$ words, $SD = 30.51$) as compared to the Black condition ($M = 30.71$ words, $SD = 22.70$), $t(56) = 0.96$, $p = .340$. \(\)
person experiencing a positive event, because both these race-valence combinations are stereotype-inconsistent.

In Study 3b, we analyzed field data on a positive lifecycle event: marriage. Specifically, we examined whether wedding articles in the *New York Times* reporting on same-sex marriages were longer than articles reporting on heterosexual marriages. Same-sex marriage violates the stereotype that same-sex relationships are less durable and loving (Peplau & Fingerhut, 2007; Testa et al., 1987). Further, they violate the prototype of marriage as involving a heterosexual couple. Examining surprised elaboration in this new domain allowed us to test for generalizability by seeing whether violations of non-racial stereotypes for positive events would also prompt elaboration.

In moving to positive domains, Studies 3a and 3b address a potential alternative mechanism of surprised elaboration: in-group favoritism. In prior studies, in-group favoritism might have led predominantly white communicators to say more about white (vs. Black or Hispanic) victims if the communicator had more compassion for in-group members, or was motivated to provide them with better service. Yet following positive events, a surprise mechanism and an in-group favoritism mechanism make opposite predictions. An in-group favoritism mechanism predicts that white communicators will say *more* about white (vs. Black or Hispanic) targets whereas a surprise mechanism predicts that white communicators will find positive events that befall white (vs. Black or Hispanic) individuals less surprising, and as a result, say *less*. So too, whereas an in-group favoritism mechanism predicts that reporters, most of whom are likely heterosexual—based on population base rates—will say more about heterosexual (vs. same-sex) weddings, a surprise mechanism predicts that reporters will say less about heterosexual (vs. same-sex) weddings.

**Study 3a: Generalizing to Positive Events with an Experiment**

In Study 3a, we tested whether people elaborate more on a white (vs. Black) teacher who was fired for sexual harassment yet elaborate more on a Black (vs. white) teacher who won an award. In both cases, we predicted that surprise would increase elaboration.

**Method**

**Participants.** We recruited participants on MTurk, following the same procedure reported in Study 2. We opened the survey to 400 participants in return for $0.40; MTurk
returned 402 respondents (53.9% female; $M_{\text{age}} = 37.27$, $SD_{\text{age}} = 12.01$). The sample was 73.0% white, 12.5% Black, 8.0% Asian, 5.5% Hispanic, and 1.0% Other.

Procedure. We used a 2 (Race: Black vs. White) × 2 (Valence: Positive vs. Negative) between-subjects design. Participants were randomized to read a scenario about a white or Black teacher that had done something either negative (sexually harassed a student; “Mr. Chonka, the teacher pictured above, was just accused of inappropriate conduct with a female student”) or positive (won an award; “Mr. Chonka, the teacher pictured above, just won a national teaching award”).

Across scenarios, the participants read information about the event and then imagined themselves as the principal at the school and wrote a memo to parents about the recent event (“Please write a memo to the parents at the high school, telling them about this recent event”). The dependent variable, as in past studies, was the number of words the participant wrote.

Results and Discussion

An ANOVA of Race × Valence revealed no main effect of Race, $F(1, 398) = 0.08$, $p = .779$, and a main effect of Valence, $F(1, 398) = 41.99$, $p < .001$. Participants wrote more about the teacher who had done something negative ($M = 43.54$ words, $SD = 25.86$) versus something positive ($M = 29.42$ words, $SD = 17.64$). In support of our hypothesis, we found a Race × Valence interaction, $F(1, 398) = 15.33$, $p < .001$. When writing a letter about the teacher guilty of sexual harassment, participants wrote more about the white teacher ($M = 47.38$ words, $SD = 26.81$) than the black teacher ($M = 39.47$ words, $SD = 24.30$), $F(1, 398) = 6.64$, $p = .010$. In contrast, when writing a letter about the teacher who won an award, participants wrote more about the black teacher ($M = 33.89$ words, $SD = 20.89$) than the white teacher ($M = 24.77$ words, $SD = 11.88$), $F(1, 398) = 8.76$, $p = .003$ (see Figure 1).
Note. Participants wrote more about negative outcomes that befell the white (vs. Black) teacher, a pattern that reversed for positive outcomes. Error bars represent one standard error.

Study 3b: The NYT Wedding Archives

We predicted that reporters would be more surprised when same-sex (vs. heterosexual) marriages take place and would therefore write longer articles about same-sex couples. To test this, we analyzed the length of all wedding articles published in the New York Times.

Method

Materials, Procedure and Measures. On October 21, 2018 we downloaded all articles from the New York Times wedding archive (https://www.nytimes.com/topic/subject/weddings-and-engagements) that had been written between January 1, 2016 and October 21, 2018. During this period, $N = 483$ wedding articles were published, 48 of them on same-sex couples (9.9%) and 435 of them on heterosexual couples (90.1%). We only included full articles (i.e., no wedding announcements) in this study.

The dataset contained the following variables, which we analyzed: (A) Elaboration. We calculated a word count for each article, which served as the primary dependent variable. (B) Sexual orientation of couple. A research assistant coded each couple as either same-sex (= 1) or heterosexual (= 0). Heterosexual couples had 1 male and 1 female. Same-sex couples had either 2 men or 2 women. (C) Date. For each article, we had the date the article was published.
Results and Discussion

On average, articles were $M = 538.92$ words ($SD = 554.73$). Supporting our hypothesis, an ANOVA of word count revealed that articles about same-sex couples ($M = 731.29$ words, $SD = 732.14$) were longer than articles about heterosexual couples ($M = 517.69$ words, $SD = 528.36$), $F(1, 481) = 6.48, p = .011, \omega^2 = .013$. This effect remained significant when controlling for publication date, $F(1, 480) = 5.95, p = .015, \omega^2 = .012$.

Across an experiment (Study 3a) and a field study (Study 3b) we find evidence that surprised elaboration occurs across positive and negative events. Testing whether surprised elaboration appeared for positive outcomes allowed us to address an alternative possibility: that communicators in Studies 1-2 elaborated more on the outcomes of white (vs. Black and Hispanic) people because they cared more about in-group members, and/or because they wanted to provide in-group members with better service. In Study 3, we did not find evidence consistent with an in-group favoritism process.

Study 4: Resource Investment in the Field

Does it matter if people elaborate more on surprising events? In Study 4, we explored the practical consequences of surprised elaboration—that is, whether people react differently to events described at greater versus lesser length. Our hypothesis was that people would want to invest more government and media resources in events accompanied by longer text. We reasoned that people would assume that a case with a longer report had been judged more important by the body that created the report (rather than inferring that the medical examiner wrote more out of surprise)—and that people would want to invest resources according to assumed importance. To test this, in Study 4 we showed each participant a random selection of reports on the unidentified bodies from Study 1b, each of which described the circumstances in which the unidentified body had been found. Participants chose whether to assign each unidentified body report high or low priority to receive government resources.

Follow-up tests demonstrated that the errors on the dependent variable (word count) was non-normally distributed. To address this we re-ran the main analysis using a heteroscedasticity-consistent standard error model. This model accounts for the normality problem, namely the fact that the errors in the data are not normally distributed. Re-running analyses using a heteroscedasticity-consistent standard error model did not change the direction or significance of the results, $B = 213.60, SE = 108.67, t(481) = 1.97, p = .049$. 
Method

Participants

We recruited participants on Prolific. Participants of any nationality were allowed to participate so long as their Prolific approval rating was at or above 50%. We opened the survey to 100 participants in return for $5.00; Prolific returned 101 respondents (49.5% female; $M_{age} = 26.58$, $SD_{age} = 8.51$).

Materials and Procedure

This study was pre-registered on AsPredicted (https://aspredicted.org/blind.php?x=2rx5xj). In the introduction, participants learned that they would be reviewing 100 selections of real-world unidentified body reports, each report describing the circumstances in which an unidentified body was found. They were told: “Please decide whether you think the government should invest in solving the case. Specifically, for each case, you will decide if you think it should be HIGH or LOW priority for receiving government funds.” Thus, for each of the 100 reports the participant decided whether the body should be high (= 1) or low (= 0) priority. Participants were told that about 50% of the reports should be assigned high priority and 50% should be assigned low priority.

Each participant viewed 100 reports of the circumstances in which an unidentified body was found; these were randomly selected from the sample of 1,190 unidentified body reports from Study 1b. Although Study 1b had 1,194 reports, we excluded the four reports that were written in Spanish. The randomly selected 100 reports were presented to participants in randomized order. Given that each of the 100 participants was randomized to review 100 unidentified body reports, each report had between 1 and 10 raters. Most of the narrative reports written by medical examiners (94.3%; 1,122/1,190) did not mention the body’s race. Rather, this information was included as a statistic, separate from the narrated circumstances in which the body was found. Thus, our research question was whether participants would prioritize cases based on the length of the report, when they were unaware of the body’s race.

---

7 We chose to test this research question in the reports of unidentified bodies (Study 1b), not missing children (Study 1a), because the missing children reports often referenced a photo and we wanted to test the effects of text length independent of pictures.
Results and Discussion

For each report, we averaged the priority ratings of all participants to create an average prioritization score between 0 and 1. On average, reports received a priority rating of $M = .50$ ($SD = .22$) suggesting that participants roughly followed the instruction to assign 50% of reports to high priority and 50% of reports to low priority. Next, we analyzed whether the word length of each report correlated with its priority rating. Supporting our hypothesis, word length correlated with the reports’ average priority ratings, $r = .15, p < 0.01, d = .30$. These results suggest that people are more interested in investing government funds in longer reports over shorter ones. (For a summary of the effects on investment decisions across Studies 4-6, see Table 3.)

Table 3

Summary of Effects on Investment Decisions across Studies 4-6

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Problem domain(s)</th>
<th>% of Participants Investing in longer reports</th>
<th>% of Participants Investing in shorter reports</th>
<th>Text Length Predicts Investment Decision</th>
<th>Cohen’s $d$</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 4</td>
<td>101</td>
<td>Unidentified Body Cases</td>
<td>—</td>
<td>—</td>
<td>$r = .15$</td>
<td>$d = 0.30$</td>
<td>—</td>
</tr>
<tr>
<td>Study 5</td>
<td>200</td>
<td>Unidentified Body and Missing Children Cases</td>
<td>64%</td>
<td>36%</td>
<td>$\chi^2(1) = 15.68, p &lt; .001$</td>
<td>$d = 0.58$</td>
<td>[0.29, 0.87]</td>
</tr>
<tr>
<td>Study 6</td>
<td>105</td>
<td>Missing Children Cases</td>
<td>78%</td>
<td>22%</td>
<td>$\chi^2(1) = 33.15, p &lt; .001$</td>
<td>$d = 1.36$</td>
<td>[0.90, 1.82]</td>
</tr>
</tbody>
</table>

Recall that in Study 1b, white (vs. Black and Hispanic) bodies received longer reports. We next tested whether the race of the individual described in each of the 1,190 reports would predict the report’s average prioritization score. Specifically, we tested whether the reports for white (vs. Black and Hispanic) bodies received higher prioritization scores, and if so, whether the greater prioritization of white (vs. Black and Hispanic) bodies was explained by their greater length. Indeed, in this sample, the reports of white bodies received higher average priority ratings ($M = .52, SD = .21$) than the reports of Black and Hispanic bodies ($M = .48, SD = .22$), $B = 0.04, SE = 0.01, t(1,188) = 3.33, p < .001, d = .19$, even though (as mentioned above) the wide majority of reports (94.3%) did not mention the body’s race in the narrative text written by the medical examiner (it was reported elsewhere). Thus, without seeing information on the body’s...
race, participants allocated more resources to white unidentified bodies than Black or Hispanic ones.

Just as in Study 1b, in this slightly smaller sample of reports (1,190 instead of 1,194) medical examiners reports of white unidentified body were longer ($M = 28.89$ words, $SD = 55.57$) than Black and Hispanic ones ($M = 21.00$ words, $SD = 20.33$), $B = 7.89$, $SE = 2.39$, $t(1,188) = 3.30$, $p = .001$, $d = .19$. When included as simultaneous predictors in a regression, both race, $B = 0.0359$, $SE = 0.0125$, $t(1187) = 2.87$, $p = .004$, and word count, $B = .0008$, $SE = .0002$, $t(1187) = 5.07$, $p < .001$, predicted prioritization. Next, we tested for mediation. Supporting our hypothesis, the greater length of white (vs. Black and Hispanic) reports explained why participants gave them higher priority, $\beta_{\text{indirect}} = 0.006$, $SE = 0.002$, 95% CI 0.004, 0.012; based on 10,000 bootstrap samples. That is, participants gave higher priority ratings to white cases over Black and Hispanic ones without having information about race. They simply did this because the white reports were longer.

Why were participants more interested in investing in longer case reports? One possibility is that the longer text suggested to participants that the case was better researched and had more valuable information that could actually be useful in solving it. Another possibility is that people use length as a heuristic—when they see longer text, they assume it is more valuable even if it does not actually contain more or better information. We tested these alternatives in Study 5.

**Study 5: Resource Investment When the Informativeness, but not Length, of Reports is Equivalent**

In Study 5, we presented participants with two reports that varied in length but contained the same information. We predicted that participants would choose to invest more resources in longer reports than shorter ones, despite longer ones having no more or better information. By varying text length, but not content, we tested whether text length was independently driving investment decisions.

We also analyzed the effect of race. That is, in addition to varying the length of the two reports, we varied whether the longer or the shorter report described a white individual or a minority member. We did this to see whether people explicitly prefer to invest resources in minority (vs. white) individuals, even if their preference to invest in lengthier texts inadvertently
leads them to do the opposite (since due to surprised elaboration, case reports for white vs.
minority members are more surprising and therefore longer).

**Method**

**Participants**

We recruited participants on MTurk, following the same procedure reported in Study 2. We opened the survey to 200 participants in return for $0.30; MTurk returned 200 respondents (42.0% female; $M_{age} = 39.92$, $SD_{age} = 13.06$). The sample was 81.5% white, 6.0% Asian, 5.5% Hispanic, 5.5% Black, and 1.5% Other.

**Procedure**

This study was pre-registered on As Predicted (https://aspredicted.org/blind.php?x=q9t2ie). We used a 2-condition (Report type: missing child vs. unidentified body) between-subjects design. Each participant saw two case reports; either two missing child reports, or two unidentified body reports. In each pair of reports, one report was short ($M = 18$ words) and one was long ($M = 56$ words). In addition, in each pair, one report was for a white individual, and one was for a Black or Hispanic individual. For the full text of all reports, see materials on OSF.

In the missing child condition, participants read about two missing children, each of which was modeled on the missing child descriptions taken from our dataset in Study 1a (e.g., “Carlos is a missing child. No one is sure of his current whereabouts. When he was last seen, he was wearing a solid red polo shirt and jeans. The red polo shirt had a collar and did not have a pattern. The jeans were standard blue jeans with pockets in the front and back, according to the report of those who last saw him.”). The second case report in the pair was shorter: it had just the essential details of the first missing child report, and altered the name and some of the facts so it would appear to be describing a different child (e.g., “Jimmy is missing. He was last seen wearing a patterned yellow t-shirt and jeans.”). The shorter report was informationally equivalent to the first report. It was shorter because it did not elaborate on details that could easily be inferred from the information provided (i.e., a pair of pants that are jeans will be blue and have pockets). We predicted that people would assume that a case with more information had been judged to be more important by the legal body that created the report (rather than seeing length as an index of stereotypicality) and that people would invest resources in the longer case according to assumed importance.
We manipulated the race of the case reports using a stereotypically white name (e.g., Allison) versus a stereotypically Black (e.g., “Shanice”) or Hispanic (e.g., “Carlos”) name, a technique for manipulating race used in prior research. Thus, each participant saw one long report and one short report. Across participants, we counterbalanced the race (white vs. Black or Hispanic) of the individual described in the long report versus the short one. To ensure that results were not driven by the details in any single case report, we randomized participants to see one of three unique (same gender) pairs of missing child reports. Upon looking at two case reports, each participant chose in which of the two case descriptions they wanted to invest (1) government and (2) media resources.

Participants in the unidentified body condition completed a similar procedure except the case descriptions were of unidentified bodies (modeled after actual descriptions from the unidentified body dataset in Study 1b). We manipulated ethnicity by adding a specification about the body’s ethnicity to each report (e.g., “white body,” “African American body,” “Hispanic body”). The follow-up questions once again asked participants to choose in which case they wanted to invest limited government and media resources.

Results and Discussion

Supporting our hypothesis, we found a main effect of narrative length on investment of government resources: more participants chose to invest government resources in the longer (64%) versus shorter (36%) case description, $\chi^2(1) = 15.68$, $p < .001$, $d = 0.58$, 95% CI = [0.29, 0.87]. This held true in both the missing child condition (68% invested in the longer report) and the body condition (60% invested in the longer report); there was no effect of condition, $\chi^2(1) = 1.39$, $p = .239$.

We also found a main effect of narrative length on investment of media resources: more participants chose to invest media resources in the longer (61%) versus shorter (40%) case description, $\chi^2(1) = 8.82$, $p = .003$, $d = 0.43$, 95% CI = [0.15, 0.71]. This held true in both the missing child condition (56% invested in the longer report) and the body condition (65% invested in the longer report); there was no effect of condition, $\chi^2(1) = 1.70$, $p = .193$ (see Figure 2).

We also found that race had a main effect on the investment of government resources: fewer participants chose to invest governmental resources in the cases of white people (43%) than Black and Hispanic people (58%), $\chi^2(1) = 4.50$, $p = .034$, $d = 0.30$, 95% CI = [0.02, 0.58].
This held true in both the missing child condition and the body condition (46% and 39% invested in the white person, respectively); there was no effect of condition, $\chi^2(1) = 1.00, p = .317$. Race also had a main effect on the investment of media resources: fewer participants chose to invest media resources in the cases of white people than Black and Hispanic people (42.0% and 58%, respectively), $\chi^2(1) = 5.12, p = .024, d = 0.32, 95\% \text{ CI} = [0.04, 0.61]$. This held true in both the missing child and the body condition (44% and 40% invested in the white person, respectively); there was no effect of condition, $\chi^2(1) = .328, p = .567$.

Taken together, these results suggest that when given race, people use it to prioritize resources away from white people, towards Black or Hispanic people. However, they also choose to invest resources in longer case reports. As a result, contrary to their explicit preference to invest in minorities, the desire to invest in longer reports leads participants—inadvertently—to want to invest more resources in struggling majority members (i.e., a white missing child, a white unidentified body) than struggling minority members (i.e., a Black missing child, a Black unidentified body).

**Figure 2**

**Resource Investment Decisions in Study 5**

![Panel 1](image1.png)  ![Panel 2](image2.png)

*Note.* Participants were more likely to choose to invest resources in cases accompanied by long (vs. short) narrative descriptions (Panel 1). They were also less likely to invest resources in cases of white (vs. Black or Hispanic) individuals (Panel 2).

**Study 6: Longer Text Without Any More Information**

The decision to invest more in a longer report could result from people observing that longer reports have greater detail and specificity. Alternatively, people might choose to invest
more in longer reports based on text length alone. To test these possibilities we asked participants in Study 6 whether they wanted to invest in a longer vs. shorter case report when the content of both reports was illegible. Thus, participants were only reacting to differences in length, not content. We predicted that participants would prefer to invest in longer versus shorter reports.

**Method**

**Participants**

We recruited participants on Prolific following the same procedure reported in Study 4. We opened the survey to 100 participants in return for $0.30; Prolific returned 105 respondents (41.9% female; $M_{age} = 28.09$, $SD_{age} = 9.67$).

**Procedure**

This study was pre-registered on As Predicted (https://aspredicted.org/blind.php?x=gx4w2m). All participants compared two hand-written case reports for missing children, both of which were illegible. One report was short and one was long (see **Figure 3**). Participants read: “In both these reports, the officer’s handwriting was not clearly legible. To give you a flavor of each report, we have transcribed the first two sentences of each report below.” These first two “transcribed” sentences were parallel across the two reports, providing the nickname of the child, where the child was from, and the child’s full name (i.e., “Arch/Gabe was last seen in Bellvue, NE/Norfolk, NE. He may go by his full name, Archie/Gabriel.” Participants were randomized to see transcribed sentences for one of three pairs of children (Arch/Gabe, Mary/Jane, Jim/John) to ensure that no one set of names or details was driving the effect.

Upon looking at two case reports, participants chose in which case to invest limited resources (“The police have to decide in which case to invest their limited resources. They only have enough resources to actively pursue one case. In which case do you think they should invest resource?”).

**Results and Discussion**

Supporting our hypothesis, we found a main effect of narrative length on investment decisions: more participants wanted to invest resources in the longer report (78%; 82/105) than the shorter one (22%; 23/105), $\chi^2(1) = 33.15$, $p < .001$, $d = 1.36$, 95% CI = [0.90, 1.82]. This suggests that text length, independent of the specificity, details, or more generally, the content of
the text, affects readers’ judgments and decisions. People show an interest in investing in longer (vs. shorter) case reports even when the longer text contain no additional information (Study 5) and even when the information in the longer text is not legible (Study 6). People appear to heuristically associate longer text with better information and to want to invest resources accordingly.

**Figure 3**

*Study 6 Manipulation*

<table>
<thead>
<tr>
<th>Report for Child #435789</th>
<th>Report for Child #435472</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRANSCRIBED: <em>Arch was last seen in Bellevue, NE. He may go by his full name, Archie.</em></td>
<td>TRANSCRIBED: <em>Gabe was last seen in Norfolk, NE. He may go by his full name, Gabriel.</em></td>
</tr>
</tbody>
</table>

*Note.* In Study 6, participants compared two illegible case reports, one longer than the other. The first two sentences of each report were “transcribed” to be legible.

**General Discussion**

Officers, medical examiners, newspaper reporters, and the public at large engage in surprised elaboration—when surprised, they elaborate. Study 1 found evidence of this using archival analysis. Officers and medical examiners wrote longer reports for stereotype-inconsistent events (i.e., a white child goes missing, a white body is found dead) than stereotype-consistent ones (i.e., a Black child goes missing, a Black body is found dead).

In follow-up experiments surprise mediated the effect of stereotype-inconsistent events on elaboration (Study 2). Communicators were more surprised by stereotype-inconsistent events which led them to write more about them. This effect generalized to positive domains, as well as to domains where greater elaboration did not favor the communicator’s in-group (Study 3). For example, in Study 3a, participants—most of whom were majority members (73.0% white)—elaborated more on Black (vs. white) award recipients. So too, in Study 3b, *New York Times* reporters—most of whom, population base rates suggest, were heterosexual—wrote more about same-sex (vs. heterosexual) weddings.
Finally, we examined implications for resource investment (Studies 4-6). Observers showed a preference for investing resources in longer case reports, likely because these seemed more important. This preference led even egalitarian-minded people to (inadvertently) prefer to invest in white victims versus Black or Hispanic victims, since the former case reports tended to be longer.

**Theoretical Implications**

Research on language and communication has often focused on the content that people communicate. Building on this research, we examined a more ancillary feature of language: the volume that people communicate. Interestingly, the study of language often obfuscates this variable. For example, software designed to study the emotionality of texts typically examines the percentage of words that are emotional: that is, the number of emotional words divided by the length of text (Pennebaker, Booth, Boyd, & Francis, 2015). While such a method is useful for accurately gauging language quality, it overlooks the volume of text communicated as a variable of interest in its own right, and how such a variable might reflect key features of human psychology. Our investigation suggests that text length is worth attending to, as it reflects key features of the communicator’s psychology, notably, the communicator’s surprise.

Indeed, elaboration follows in the wake of surprise: communicators say more about events they find surprising. While past research has shown that surprise increases the tendency to explain events (Hastie, 1984; Sekaquaptewa, Espinoza, Thompson, et al., 2003; Von Hippel et al., 1997), we found that surprised elaboration occurred even in the absence of explanations. For example, reports of unidentified bodies in Study 1 were longer when they described stereotype-inconsistent cases, but they did not contain more explanations. In general, these reports contained few explanations given that medical examiners are typically sharing facts about unidentified bodies, not explaining “why” an event occurred. While the missing child reports of white (vs. Black and Hispanic) children did contain more explanations, only detail and specificity accounted for the effect of race on elaboration (and, once again, few reports offered any explanations at all).

In addition to indexing the psychology of the communicator, we found text length to have an important effect on the receiver. The primary purpose of communication is to be consumed—people write so others will read what they say. Thus, key to the full study of language is to understand the effect that language has on the people reading it. In Studies 4-6, we explored how...
observers reacted to long (vs. short) text. We found that the length of unidentified body reports influenced investment decisions—on average, over a large sample of unidentified body reports, longer texts garnered more interest in investments (Study 4). In these real-world texts, it is possible that participants wanted to invest in longer texts because they perceived these texts to have greater detail and specificity. However, follow-up experiments demonstrated that when the detail and specificity of longer and shorter texts were equated (Study 5), or hidden entirely (Study 6), observers continued to show an interest in investing in longer texts over shorter ones.

Our findings contribute to the stereotype literature by identifying a new, seemingly innocuous consequence of stereotypes: surprised elaboration. Since stereotypes inform what people expect, they also affect surprise, and thereby, the amount that people communicate. Moreover, surprised elaboration not only reflects stereotypes but perpetuate downstream inequality—by leading observers to invest more in longer (vs. shorter) negative narratives of white (vs. Black or Hispanic) victims. Of course, for positive events, surprised elaboration can have the reverse effect, leading people to draw more attention to positive events involving Black and Hispanic (vs. white) individuals, because again, the former are more surprising. Nevertheless, our results suggest that among those in dire circumstances, white people are more likely to get the help they need. Given that receiving help and resources in negative circumstances is more essential than receiving help and resources in positive circumstances, there is reason to believe that the asymmetry does not cancel itself out. The relative lack of help given to Black and Hispanic individuals in hard straits is arguably more damaging than receiving extra help in good times.

**Alternative Explanations and Theories**

In the current investigation, we brought evidence for surprised elaboration. This model proposes that stereotype-inconsistent events generate more communication than stereotype-consistent events because the former are more surprising.

Another possibility is that communicators in our studies—many of whom were white—elaborated more on stereotype-inconsistent events because doing so was a service to the in-group. Indeed, in Study 1, professionals who, based on population statistics, can be assumed to have been primarily white (Police officers, n.d.)—wrote longer reports on missing white (vs. minority) children, or unidentified white (vs. minority) bodies. In doing so, they performed a service to their in-group. Nevertheless, in-group favoritism appears an insufficient explanation
for other studies. For example, in Study 3a, white communicators elaborated more on Black (vs. white) teachers who won awards, a disservice to the in-group. In-group favoritism is also an insufficient explanation in Study 3b since reporters, most of whom were probably heterosexual—again, based on population base rates—elaborated more on out-group (same-sex) than in-group (heterosexual) weddings. Thus, while in-group favoritism might be an alternative explanation for some of the results in the paper, it is an unlikely explanation for all of them.

Another alternative possibility is that communicators were more motivated to elaborate not because they were personally surprised, but because they anticipated the audience would be. Indeed, an event that violates common stereotypes violates the expectations of both the communicator and the audience. While we believe that both parties are surprised by stereotype-inconsistent events, to test whether the surprise of the communicator drove the effect, we ran several experiments in which communicators wrote communications that went to no one except the researcher, thus documenting that motivated elaboration occurred even when the role of the audience was minimized.

Finally, we consider other reasons our communicators were surprised. Our theory is that violated stereotypes caused the surprise and hence, prompted elaboration. An alternative possibility is that peoples’ personal experience informed when they felt surprised. For experts, there may often be a difference between an event that violates common stereotypes and an event that violates their personal experiences. For example, if a reporter for the New York Times is always assigned to cover same-sex weddings, after covering dozens of such weddings they will no longer feel surprised. While expert scenarios like these likely exist in the world, we do not believe that surprise based on personal experience drove the observed effects in our studies. Indeed, the New York Times does not have reporters who specialize in covering same-sex weddings. So too, the police officers, sheriffs, and other government officials who wrote up the reports on missing children and unidentified bodies in Study 1 were not missing child or unidentified body specialists (Bartlett, 2011; Kelly, 2020). Writing up rare events like these is a small part of their jobs. Since communicators in our studies were discussing events with which they themselves had only limited personal experience, we expect that their surprise was guided by violations of societal expectations rather than personal experience.
Future Directions

Future research might continue to examine whether surprised elaboration is affected by the communicator’s race. Our preliminary evidence suggests it is not. We documented surprised elaboration across both white and Black participants. We predicted surprised elaboration would generalize across participants of both races given that common stereotypes are just that—common. They are held by majority members and minority members (Eberhardt, 2020). One of the reasons stereotype threat is so powerful is because often, marginalized group members have internalized negative stereotypes, and as a result, hold negative expectations for themselves.

The policy consequences of motivated elaboration also deserve further study. For example, if people infer that a missing child with a longer case report deserves more resources, do they also infer that a human life with a longer description has a larger life insurance value, or that a gestating fetus that is described in lengthier terms is more viable? Just as the length of a report led observers to invest more resources, text length might lead observers to judge the target as more important, valuable, or alive, with consequential effects.

Finally, insofar as peoples’ surprise in the face of stereotype-inconsistent events prompts elaboration, future researchers might explore narrative elaboration as a new tool for revealing authors’ stereotypes. Stereotypes are notoriously difficult to measure, since people are often not willing to disclose their prejudices, and even if they are, they may not be aware of them (Fazio et al., 1995; Greenwald & Banaji, 1995). Implicit measures, such as the IAT, attempt to circumvent these problems, but task measures have limitations: they can only be used if people choose to sign in and complete a research task, and it is up to the test-taker whether to consider or discount the test results. Word count, by contrast, can be detected in any publicly-available narrative text. Unbeknownst to communicators, the degree to which they elaborate reflects their surprise, and thus, whether an event in question violated a held stereotype. Future research might explore whether surprised elaboration can be used as a new measurement tool for identifying the stereotypes and biases of the communicator.

Conclusion

Narrative elaboration is a key aspect of communication. Every author, high school student, and newspaper reporter knows that word count is king. Readers glancing at a page in a newspaper intuit that the longer article is probably more attention-worthy, just as students know that submitting an essay that is too short will arouse suspicion that no effort was expended. This
Despite the fact that great writers often advise us to do the opposite, Mark Twain once told a correspondent, “If I had had more time I would have written you a shorter letter” (Twain, 1871), a sentiment echoed in Strunk and White’s famous admonishment: “omit needless words” (Strunk & White, 1979).

In the current paper, we do not offer writerly advice. We study text length from a psychological perspective. While writers are often advised to think about word count for stylistic reasons, we point to the psychology of longer texts—what prompts communicators to elaborate in the first place, and how lengthier text influences the reader.
References


Ireland. https://doi.org/10.1145/2660460.2660469


