Generic language in scientific communication

Jasmine M. DeJesus[a,1], Maureen A. Callanan[b], Graciela Solis[c], and Susan A. Gelman[d,e,1]

[a]Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC 27412; [b]Department of Psychology, University of California, Santa Cruz, CA 95064; [c]Department of Psychology, Loyola University, Chicago, IL 60660; [d]Department of Psychology, University of Michigan, Ann Arbor, MI 48109-1043; and [e]Department of Linguistics, University of Michigan, Ann Arbor, MI 48109-1043

Scientific communication poses a challenge: To clearly highlight key conclusions and implications while fully acknowledging the limitations of the evidence. Although these goals are in principle compatible, the goal of conveying complex and variable data may compete with reporting results in a digestible form that fits (increasingly) limited publication formats. As a result, authors' choices may favor clarity over complexity. For example, generic language (e.g., "Introverts and extraverts require different learning environments") may mislead by implying general, timeless conclusions while glossing over exceptions and variability. Using generic language is especially problematic if authors overgeneralize from small or unrepresentative samples (e.g., exclusively Western, middle-class). We present 4 studies examining the use and implications of generic language in psychology research articles. Study 1, a text analysis of 1,149 psychology articles published in 11 journals in 2015 and 2016, examined the use of generics in titles, research highlights, and abstracts. We found that generics were ubiquitously used to convey results (89% of articles included at least 1 generic), despite that most articles made no mention of sample demographics. Generics appeared more frequently in shorter units of the paper (i.e., highlights more than abstracts), and generics were not associated with sample size. Studies 2 to 4 (n = 1,578) found that readers judged results expressed with generic language to be more important and generalizable than findings expressed with nongeneric language. We highlight potential unintended consequences of language choice in scientific communication, as well as what these choices reveal about how scientists think about their data.

generic language | scientific communication | diversity | metascience | psychological research

Recent discussions of scientific practices in the social sciences reveal 2 themes that appear to be at cross-purposes. On the one hand, there is increasing concern about samples that are limited in scope and generalizability. Research in psychology often relies on samples from WEIRD (Western, educated, industrialized, rich, and democratic) societies that are unrepresentative of people the world over (1–3). Furthermore, sample sizes are often underpowered (4–6), at times leading to conclusions that do not hold up with larger-sized replications or meta-analyses (7). At the same time, scientists are increasingly encouraged to describe their work in an accessible manner, to reach out to broad audiences, and to make bold, interesting claims about the wider implications of their findings (tweets, TED talks, and so forth) (8). These trends are reflected in the introduction of new condensed formats (such as research highlights) and metrics that focus on popularity and uptake (journal impact factors, H-indexes, AltMetrics). Together, these 2 themes present a challenge for scientific writing: To communicate key findings in an accessible and concise manner, while fully and responsibly acknowledging the variability and limits of the evidence. Precision may be sacrificed when attempting to reach a broader audience, and diversity in findings may be ignored. For example, university press releases contain more exaggerated advice, exaggerated causal claims, and exaggerated inference to humans from animal research than the original peer-reviewed

journal articles they summarize (9). Does scientific writing itself also fall prey to similar tendencies?

Here we examine one way that authors of peer-reviewed scientific reports make choices that favor brevity over precision. We focus on the use of generalizing claims based on limited evidence: Broad claims about "infants," "Whites," "Millennials," "women," or "adolescents with seasonal affective disorder." Examples include: "Whites and Blacks disagree about how well Whites understand racial experiences," "Americans overestimate social class mobility," "animal, but not human, faces engage the distributed face network in adolescents with autism," or "women view high-level positions as equally attainable as men do, but less desirable." This tendency has been recognized in science writing informally for decades. Oyama (10, 11) argued that when reasoning about human nature, theorists often conflate incidence (relative frequency or probability of a trait) with essence ("a hidden truth, rooted in the past and already there"). Thus, people often pose general questions—such as "Are people at their core aggressive or peaceful, selfish or altruistic, rational or irrational?"—rather than viewing behavior within a dynamic developmental system (12). Siegler likewise suggests that developmental psychologists tend to focus on questions such as, "[W]hat is the nature of 5-year-olds' thinking, and how does it differ from the thinking of 8-year-olds?", treating each age group as a uniform and static entity, rather than considering the variability within each age group and processes of continuous change (12). Similarly, Barrett (13) suggests that scientists may unwittingly incorporate essentialist assumptions into their theorizing, ignoring

Significance

There is increasing recognition that research samples in psychology are limited in size, diversity, and generalizability. However, because scientists are encouraged to reach broad audiences, we hypothesized that scientific writing may sacrifice precision in favor of bolder claims. We focused on generic statements ("Introverts and extraverts require different learning environments"), which imply broad, timeless conclusions while ignoring variability. In an analysis of 1,149 psychology articles, 89% described results using generics, yet 73% made no mention of participants' race. Online workers and undergraduate students (n = 1,578) judged findings expressed with generic language more important than findings expressed with nongeneric language. These findings provide a window onto scientists' views of sampling, and highlight consequences of language choice in scientific communication.

or downplaying variation in favor of averages or central tendencies, resulting in reports that extend beyond the evidence, imply unwarranted uniformity and universality, and downplay variability and contextual influences.

A tendency to generalize broadly from samples and gloss over variation with law-like statements could be especially problematic when combined with the lack of diversity of participants in most social science studies, noted earlier (14). Rogoff (15) and Gutiérrez and Rogoff (16) warn of overly general claims (e.g., "children do such and such") that assume "a timeless truth" and generalize too quickly to other populations without evidence. These statements hide that study participants typically include a limited and unrepresentative group of individuals and imply that this group is the norm and that their behaviors, attitudes, and perceptions are universal. Gutiérrez and Rogoff recommend replacing such broad claims with past-tense statements that convey what was observed in a given situation ("the children did such and such"), noting, "Only when there is a sufficient body of research with different people under varying circumstances would more general statements be justified" (16). Generalizing from WEIRD samples also encourages deficit thinking; when participants from non-WEIRD samples perform differently, this is often described as abnormal or problematic (15, 17–19).

The most common means of expressing generalizations is via generic language statements regarding categories, such as "dogs are 4-legged" or "tigers have stripes" (20–27). Generics make broad claims about a category as a whole, as distinct from individuals, without reference to frequencies, probabilities, or statistical distributions. Thus, the generic claim "people use the availability heuristic" implies a broader truth, in contrast to: "the people in this study used the availability heuristic" (describing the behavior of individuals within a sample), "most people in this study used the availability heuristic" (describing a probability), or "60% of people in this sample used the availability heuristic" (describing a statistical likelihood). Generics have been observed across all languages that have been studied (20), and they are understood and produced early in childhood—by about 2.5 y of age (27, 28)—earlier than the acquisition of quantifiers, such as "all," "some," or "most" (29–31). They appear to be a default mode of generalization, with quantified statements often misremembered as generic (32).

Generics have important semantic implications. They gloss over exceptions and variability. For example, one can say "birds fly" even though penguins and ostriches do not fly, one can say "lions have manes" even though only male lions have manes, and one can say "mosquitoes carry the West Nile virus" even though fewer than 1% of mosquitoes do so (22, 25). Moreover, generics regarding animal kinds exhibit an asymmetry: People need relatively little evidence to make a generic claim, but generic claims are interpreted as applying broadly to the category (33–35). This is not true of quantifiers (such as "all" or "most"), for which evidence and interpretation match up precisely. Generics are resistant to counter evidence; for example, the claim "introverts and extraverts require different learning environments" (36) is not disconfirmed by an instance of an introvert and an extravert who require the same learning environment. Furthermore, generics imply that a feature is conceptually central (37–40) and can lead to higher rates of essentialism (41). Generics may also imply that a statement is normatively correct or ideal (39, 40, 42, 43). For all of these reasons, expressing a finding generically may lead readers to think that a scientific result is especially important (because it is conceptually central and universal), robust (because it is normatively correct), or generalizable (because it downplays variability and exceptions). To this point, however, the use and interpretation of generic language in science writing has been unexplored.

The goal of the present paper is to conduct a systematic examination of the frequency of generic language in reporting results of psychology publications, to examine competing hypotheses regarding their contexts of use, and to examine how lay readers interpret such statements. Because generics are a default mode of generalization and the most common means of expressing generalizations in natural language, including for variable tendencies that do not uniformly apply in all instances, we anticipated that they would be widely used to express results in the psychological sciences. Because generics gloss over exceptions and imply that variation does not exist, we hypothesized that they may more likely be found in articles that choose not to report how individuals in their samples vary from one another. Similarly, because generics are universalizing and imply that findings are broadly true regardless of time and place, we hypothesized that findings expressed with generic language would be interpreted as more important and conclusive by lay readers, even if there was little evidence linking the use of generic language by authors and the evidentiary basis for their findings.

Study 1: Generic Language in Published Psychology Articles

The goal of study 1 was to assess the frequency of generic language in a corpus of scientific papers in psychology. We focused on psychology because of the participant sampling issues that are acute in the social sciences, including overreliance on WEIRD populations and small sample sizes. Journals were selected to range broadly across subfields (biological, clinical, cognitive, developmental, social), to have relatively high impact factors, and to include research highlights or similar short summaries as well as abstracts. Analyses were restricted to titles, highlights, and abstracts, because these tend to be the most read components (44) and may be critical in determining whether an article is sent for review or read further. Furthermore, researchers may resort to generics, especially when word counts are restricted, which is true for all 3 of these components.

We included all articles published in 2015 and 2016 in 11 journals, excluding articles that did not report results with human participants or did not provide highlights, resulting in 1,149 journal articles (see *SI Appendix*, Table S1 for exclusions and *SI Appendix*, Table S2 for journal details). For each article, we coded each complete sentence in the title, highlights, and abstract that pertained to the results of the study. These were coded as either generic sentences (in 1 of 3 forms; see below) or nongeneric sentences. Generics were general, timeless claims regarding categories or abstract or idealized concepts (e.g., "infants," "sleep," "the brain"), and were coded as 1 of 3 types: Bare, framed, or hedged.

"Bare" is a generic sentence that is unqualified and not linked to any particular study, such as "infants make inferences about social categories" or "adolescent earthquake survivors' [sic] show increased prefrontal cortex activation to masked earthquake images as adults."

"Framed" is a generic that is unqualified, but is framed as a conclusion from the particular study that was conducted, such as "Moreover, *the present study found that* dysphorics show an altered behavioral response to punishment" or "*We show that* control separately influences perceptions of intention and causation" (emphases added).

"Hedged" is a generic that is qualified by a phrase, such as "*These results suggest that* leaders emerge because they are able to say the right things at the right time" or "Thus, sleep *appears to* selectively affect the brain's prediction and error detection systems," or a qualifying adverb (e.g., "perhaps") or auxiliary verb (e.g., "may"), such as "Mapping time words to perceived durations *may* require learning formal definitions" (emphases added).

Sentences were coded as nongeneric when they did not make a general claim, and typically were worded in the past tense (e.g., "Mortality salience increased self-uncertainty when self-esteem was not enhanced"; "Differences in telomere length were not due to general social relationship deficits"). Sentence fragments that omitted the main verb (e.g., "implicit fear and effort-related

PSYCHOLOGICAL AND COGNITIVE SCIENCES

cardiac response") were excluded from our analyses because they were unable to reveal whether the claim was generic (e.g., "implicit fear determines effort-related cardiac response") or not (e.g., "implicit fear was associated with effort related cardiac response"). Sentences were excluded from the coding of the highlights or abstracts if they referred to previous research or methodology, to restrict comparisons to the research findings on which generic claims were based. We considered all titles to pertain to the results, because titles often alluded to an undifferentiated combination of results, research question, and/or method (e.g., "eye movements reveal memory processes during similarity- and rule-based decision making").

Generics were hypothesized to be prevalent across content areas of psychology, given the frequency of generics in natural language and their utility for making broad claims in the face of variable data. We begin by reporting the frequency of generic language in our sample as baseline data. We then examine whether generics are linked to features of the articles, including the format length (i.e., Are generics more common in shorter or longer formats?) and sample features (i.e., Are generics more common in articles that included larger samples?). Generics were also hypothesized to be independent of the evidentiary basis for the claim (i.e., generics were not expected to be more frequent in articles with more participants or participants from more diverse backgrounds), given that they are a default mode of generalization. Finally, generics were hypothesized to be more frequent in shorter than longer formats (titles and highlights vs. abstracts), given that generics entail the absence of specification and thus are typically shorter than nongeneric claims (22). A corollary to this prediction was that, when generics were used, shorter formats (e.g., titles and highlights vs. abstracts) were expected to elicit higher rates of unqualified (bare) generics, compared to framed and hedged generics, given that bare generics are shortest.

Results

Frequency of Generic Language. This corpus of 1,149 articles included 13,978 codable elements (i.e., sentences reporting results): 358 codable titles, 4,409 codable sentences from highlights and other short summaries, and 9,176 codable sentences from abstracts. Generic language was prevalent in these articles: 89% of the articles (1,025 of 1,149) included at least 1 generic sentence. As a percentage of codable sentences, generics were most frequent for shorter elements: Titles (87%) > highlights (40%) > abstracts (16%) (Fig. 1). Note, however, that most titles (69%)

were uncodable (i.e., missing a main verb), whereas 99% of highlights and abstracts were codable.

Associations between Generic Language and Format Length. To examine whether generics were more common in shorter article formats and whether generic language use was associated with journal impact factor, a multilevel model was performed on the proportion of bare, framed, or hedged generics (as a function of the number of codable sentences) in each component as the outcome variable (Table 1). Article component (title and highlight vs. abstract) and the impact factor for the journal in the year the article was published were entered as predictors. Compared to abstracts (mean = 0.16), highlights (mean = 0.40, $b = 0.25$, SE = 0.01, $t = 21.70$, $P < 0.001$) and titles (mean = 0.87, $b = 0.71$, SE = 0.02, $t = 42.26$, $P < 0.001$) had relatively more generics. No significant effect of journal impact factor was observed. To directly compare the nonabstract formats to one another, the model was repeated with highlights as the reference category; titles had a significantly higher proportion of generics than highlights ($b = 0.47$, SE = 0.02, $t = 27.54$, $P < 0.001$).

Because many titles were uncodable (therefore had 0 in the denominator of the proportion), the above analysis was restricted to the 358 articles in which all components had at least 1 codable sentence. In order to include all articles ($n = 1,149$), we reran the analysis on just highlights and abstracts and observed the same effect of component: Highlights (mean = 0.40) were more likely to include generics than abstracts (mean = 0.16), $b = 0.25$, SE = 0.01, $t = 23.02$, $P < 0.001$. Again, no significant effect of journal impact factor was observed, $P = 0.964$.

In addition to the overall usage of generic language, we were also interested in the varieties of generic language that were employed. Bare generics were unqualified (compared to generic statements that were qualified in some way, either hedged or framed within the context of a study) and so the most starkly generalizing. Because bare generics have fewer linguistic elements, we predicted that such forms would be more common in shorter formats. To test this hypothesis, a χ^2 test was performed on sentences that were coded as generic (bare, framed, or hedged). A significant association was observed: $\chi^2(4) = 1222.53$, $P < 0.001$ (Table 2). Within the sentences coded as generic, bare generics were more common in shorter formats, accounting for 98% of titles, 70% of highlights, and 15% of abstracts. Inversely, framed and hedged generics were more common in longer formats.
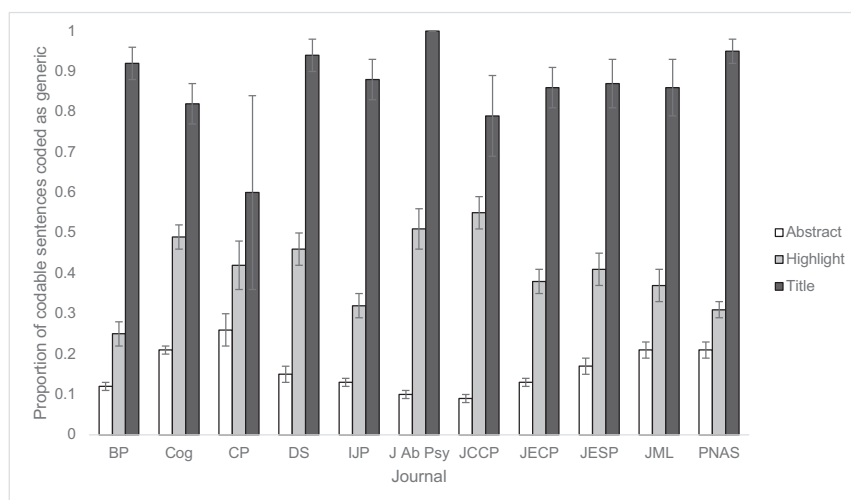
Fig. 1. Study 1: The proportion of sentences in titles, highlights, and abstracts coded as generic (number of generic sentences divided by the number of codable sentences for that component, to derive a percentage), separately by journal.

Table 1. Study 1: Regression table comparing the prevalence of generic language by article component (titles and highlights vs. abstracts; highlights vs. title) and the journal impact factor for the year of publication

Predictor	Estimate	SE	t value	P value
(Intercept)	0.14	0.03	4.87	<0.001
Component				
Title vs. abstract	0.71	0.02	42.26	<0.001
Highlights vs. abstract	0.25	0.01	21.70	<0.001
Title vs. highlights	0.47	0.02	27.54	<0.001
Impact factor	<0.01	< 0.01	0.42	0.680

Associations between Generic Language and Sample Features. In addition to coding generic language, we also coded features of the participant samples in each article, with the goal of examining whether generic language use was related to the evidentiary basis of the articles (sample size [the number of participants] and sample diversity [variation in participants' racial/ethnic, socioeconomic, and language backgrounds]), as well as journal impact factor. These data are reported separately by journal in *SI Appendix* (*SI Appendix*, Table S3). Notably, most articles did not specify participants' race/ethnicity (73%), socioeconomic status (79%), or language background (74%). Because most articles did not even report this information, we were unable to relate generic language to sample diversity. Therefore, we instead conducted an analysis to test whether generics were more frequently used for articles that glossed over participant demographics, by comparing studies that reported participant background to studies that did not specify this information.

To do so, a multilevel linear regression was conducted on the proportion of generic sentences (bare, framed, and hedged generics divided by the number of codable sentences per article), with number of participants entered as a continuous predictor, the country of recruitment, participant race, participant socioeconomic status (SES), and participant language background entered as categorical predictors (with "unspecified" set as the reference category), and journal impact factor entered as a nested predictor. Two significant predictors emerged: Whether the article specified the participants' SES ($b = -0.04$, SE = 0.01, $t = -3.29$, $P = 0.001$) and whether the article specified the participants' language background ($b = 0.02$, SE = 0.01, $t = 1.99$, $P = 0.046$) (Table 3). Articles in which participants' socioeconomic background was not specified ($n = 906$) included more generics (mean = 0.25, SE = 0.01) than articles that specified some aspect of participant SES ($n = 243$; mean = 0.20, SE = 0.01). In contrast, articles in which the participants' language background was mentioned ($n = 296$) included more generics (mean = 0.26, SE = 0.01) than articles that did not specify language background ($n = 853$; mean = 0.24, SE = 0.01).

To summarize the results of study 1: Generic language was frequently used to characterize psychological results across a broad range of highly ranked journals. This practice was especially common in shorter article formats, such as titles and highlights, which provide less opportunity to communicate nuances in the findings or limitations of the work. Generic sentences were also less likely to be framed or hedged in shorter article formats. Generic use was unrelated to the evidentiary basis of the claim (as measured by the features of the sample coded from the articles): Articles that recruited a larger sample were not more likely to include generics than articles that reported smaller samples. For the 2 features that were associated with generic usage (reporting of SES and language background), generic language use was inconsistently related to authors' reporting of sample features: Articles that did not report the socioeconomic background of the participants were more likely

to use generic language, whereas those that did report the language background were more likely to use generic language. Reporting on these factors may have qualitatively different roots. Authors who report SES might be more sensitive to the constraints of their findings in their reporting (and therefore be less likely to use generics). In contrast, papers that report language are often specifically asking questions pertaining to language and comparing groups; this tendency was much more common in the *Journal of Memory and Language* (an outlier in this regard; 72% reported language) than other journals (language reporting ranged from 8 to 37%). It is possible that these comparisons could elicit more generic language. These hypotheses are speculative, but would be interesting directions for future study.

Studies 2 to 4: Readers' Inferences About Scientific Findings

Study 1 highlighted the ubiquity of generic language in published psychology articles. Studies 2 to 4 examined if and how generic versus nongeneric language influenced how summaries of research findings were interpreted by nonscientists, primarily samples of online workers, with one sample of undergraduate students enrolled in introductory psychology. Study 2 focused on generics versus simple past-tense nongenerics, study 3 examined the implications of multiple linguistic cues to nongenericity, and study 4 provided a more sensitive assessment of distinctions between generic and nongeneric wording. Each experiment within studies 2 to 4 is fully described in the *SI Appendix*, Tables S4–S9 and is summarized here. Study 2 was approved by the University of Michigan Institutional Review Board, "Language in scientific findings," HUM00131970; the protocol was determined to be exempt from ongoing institutional review board review and covered all subsequent studies, including studies 3a to 3d and 4a to 4b. Studies 3a to 3d and 4a to 4b were also approved by the University of North Carolina, Greensboro Institutional Review Board: "Language in scientific findings," 1-0332.

Study 2: Judgments of Generics vs. Simple Past-Tense Nongenerics

Participants were Amazon Mechanical Turk workers in the United States ($n = 416$) (see *SI Appendix*, Table S4 for complete demographic information). We manipulated the content of the summaries by selecting 60 titles from the study 1 corpus: 10 each from 5 different content areas of psychology (biological, clinical, cognitive, developmental, social) and PNAS. Hedged, framed, and nongeneric versions of each title were created from the bare generic version to control for article content across participants. Nongenerics were minimally cued by simply changing the tense of the verb in the bare generic (e.g., "group discussion improves lie detection" [bare generic] vs. "group discussion improved lie detection" [past-tense nongeneric]). Hedged and framed versions were created by adding elements (such as "this study suggests" or "this study confirms") to the bare generic. Titles were described as "a brief summary of different research projects" and participants were randomly assigned to complete 1 of 4 test questions for each summary: Importance ("How important do you think the findings of this research project are?"), generalizability ("What percentage of people in the world today would show the effect described in this research project?"), sample size ("How many people participated in this research project?"), and

Table 2. Study 1: Number of lines coded as bare, framed, and hedged generic (excludes uncodable and nongeneric lines)

Article component	Bare	Framed	Hedged
Abstract	200	512	643
Highlights	1,171	204	305
Title	306	0	7

Table 3. Study 1: Regression table testing for associations between generic language use and the features of individual articles

Predictor	Estimate	SE	t value	P value
(Intercept)	0.25	0.01	22.16	<0.001
Impact factor	<0.01	<0.01	0.73	0.464
No. of participants	<0.01	<0.01	0.48	0.635
Test location (vs. unspecified)				
United States only	<−0.01	0.01	<−0.01	0.999
Not just United States	<0.01	0.01	0.31	0.755
Participant race (specified vs. unspecified)	<−0.01	0.01	−1.02	0.310
Participant SES (specified vs. unspecified)	−0.04	0.01	−3.29	0.001
Participant language (specified vs. unspecified)	0.02	0.01	1.99	0.046

diversity ("How likely do you think it is that this finding would extend to people from diverse backgrounds?"). Our primary research question was whether participants' judgments varied by the type of language used to describe findings. Multiple regression models were performed, one per question, with generic language and content area entered as predictors (*SI Appendix*, Table S4).

Generic Language. Participants were sensitive to the generic language manipulation when asked to rate the importance of each summary (see *SI Appendix*, Table S5 for full tables from an ordinal regression model). Bare (mean = 4.12, $b = 0.15$, $P = 0.018$), framed (mean = 4.19, $b = 0.20$, $P = 0.002$), and hedged generics (mean = 4.18, $b = 0.19$, $P = 0.002$) were all rated as more important on a 1 to 7 scale than nongeneric summaries (mean = 4.03). Under some circumstances, participants also considered generic language when rating generalizability and sample size: Participants rated framed generics (mean = 55.70%, $b = 0.24$, $P = 0.023$) as generalizable to a larger percentage of people than nongenerics (mean = 53.47%) and rated hedged generics (mean = 3.94, $b = 0.14$, $P = 0.033$) as being drawn from larger samples (rated on a 1 to 7 scale) than nongenerics (mean = 3.86). There were no effects of generic language when judging sample diversity.

Content Area. The content area of the summaries affected participants' judgments on all 4 test questions (*SI Appendix*, Table S5). For example, clinical summaries were rated as more important and having larger samples but less generalizable and extending to less diverse samples than PNAS summaries.

To summarize, participants in study 2 judged findings expressed with generic language as more important (and at times more generalizable and drawing from a larger sample) than findings expressed with nongeneric language, despite a subtle language manipulation (simply varying verb tense, for example, from "improves" [generic] to "improved" [nongeneric]). Nonetheless, the effects were small, so we conducted 2 additional studies to understand more fully the conditions under which generic language affects readers' interpretations of research summaries.

Studies 3a to 3d: Multiple Cues to Nongenericity

In studies 3a to 3d, we sought to replicate and extend the findings of study 2 by testing a broader range of linguistic cues to nongenericity. Language can signal nongenericity in a rich variety of ways, and most prior research has provided a starker contrast between generic and nongeneric than was provided in study 2 (28–35). For example, whereas study 2 only manipulated tense, in the published literature, nongenerics typically manipulated the noun phrase itself (e.g., "This X…", "Some Xs…"), thus drawing attention more explicitly to the limited scope of the generalization. Thus, studies 3a to 3d aim to examine the semantic contrast when nongenerics are cued in a variety of ways.

Each experiment is fully described in *SI Appendix* (*SI Appendix*, Tables S6 and S7) and summarized here.

Study 3a ($n = 74$) provided 3 kinds of sentences: Bare generics (e.g., "People with dysphoria *are* less sensitive to positive information in the environment"); nongenerics signaled by past tense alone (as in study 2; for example, "People with dysphoria *were* less sensitive to positive information in the environment"); and nongenerics signaled with multiple cues, including past tense, an explicit qualifier, and the word "some" inserted into the subject noun phrase (e.g., "*Some* people with dysphoria *were* less sensitive to positive information in the environment, *under certain circumstances*") (emphases added here only for clarification; no words were italicized in the study). For each summary, participants were asked to rate the finding's importance, how much they would want to draw conclusions from the finding, and whether the finding would generalize within and outside the United States. Across all measures, participants rated the bare generics (mean = 4.52) as more important than the multicue nongenerics (mean = 3.85, $b = −0.77$, $P < 0.001$), but in contrast to study 2, bare generics were rated as equivalent to simple past-tense nongenerics (mean = 4.54, $b = 0.06$, $P = 0.511$). We will revisit this null result in studies 4a and 4b, and discuss what it means in *General Discussion*.

Studies 3b and 3c ($n = 382$) aimed to replicate study 3a, but each participant answered only 1 type of question (importance or conclusiveness). Participants were shown bare generics, past-tense nongenerics, and multicue nongenerics and were asked to either rate the importance of the finding ($n = 195$) or how much they would want to draw conclusions from the summary ($n = 187$). Study 3b recruited participants from Amazon's Mechanical Turk ($n = 264$) and 3c recruited participants from the University of Michigan undergraduate psychology participant pool ($n = 118$). In both studies, participants rated bare generics (mean = 4.50) more highly than multicue nongenerics (mean = 3.92, $Ps < 0.001$); again, however, no difference was observed between bare generics and simple past-tense nongenerics (mean = 4.52, $Ps > 0.5$). Additionally, there were no differences between the 2 samples ($Ps > 0.4$) and no interaction between participant sample and generic language ($P = 0.645$).

Study 3d ($n = 299$) provided a more fine-grained analysis of the point at which participants' ratings of nongeneric statements differed from generics. As in study 3c, participants were shown bare generics, simple past-tense nongenerics, and multicue nongenerics. They also received 2 additional forms of nongenerics, each with a subset of the cues from the multicue version: Qualified nongenerics (e.g., "People with dysphoria *were* less sensitive to positive information in the environment, *under certain circumstances*") and "some" nongenerics (e.g., "*Some* people with dysphoria *were* less sensitive to positive information in the environment"). Participants rated bare generics (mean = 4.51) more highly than multicue nongenerics (mean = 4.19, $Ps < 0.005$) and qualified nongenerics (mean = 4.25, $Ps < 0.05$). For "some" nongenerics and past-tense

nongenerics, comparisons to bare generics varied by question (importance: no differences; conclude: bare higher than "some" and lower than past-tense nongeneric) (*SI Appendix*, Fig. S2 and Table S7).

Overall, studies 3a to 3d provide further evidence that online and undergraduate student samples used generic language as a cue to evaluate the importance of research findings. Across 4 experiments, we found a persistent advantage for generic language as compared to nongenerics expressed with multiple cues (tense plus qualifier, tense plus quantified noun phrase ["some Xs"], or tense plus both qualifier and quantified noun phrase). Nonetheless, in contrast to study 2, participants did not rate simple past-tense generics differently from bare generics. In study 4, we directly contrasted bare generics with other forms to examine participants' sensitivity to these subtle linguistic differences.

Studies 4a and 4b: Direct Language Comparison

Studies 4a and 4b provided a more sensitive assessment by providing participants with directly contrasting statements varying only in wording. Each trial included a bare generic paired with a sentence expressing identical content but in a different form. In study 4a, all trials compared a bare generic with a past-tense nongeneric; in study 4b, the comparison to the bare generic was either a framed generic, past-tense nongeneric, qualified nongeneric, "some" nongeneric, or multicue nongeneric (with equal numbers of each type of comparison) (see *SI Appendix*, Table S8 for participant demographics, and *SI Appendix*, Fig. S3 and Table S9 for results). MTurk participants ($n = 407$) were asked which of the 2 summaries was more important ($n = 206$) or which they would rather draw conclusions from ($n = 201$), each rated on a 1 to 7 scale. In study 4a, participants rated bare generics as more important than simple past-tense nongenerics describing the same content [mean $= 4.42$, $t(103) = 4.07$, $P < 0.001$], and in study 4b, they judged bare generics as more important than all other nongeneric alternatives ($Ps < 0.007$), but as less important than framed generics [mean $= 3.32$, $t(101) = -5.61$, $P < 0.001$]. We thus replicated that bare generics were viewed as more important than nongenerics, including even the most subtle form (only differing in whether sentence was phrased with present or past tense verbs), but also that anchoring a general conclusion to scientific research by means of a framed generic (e.g., "*This study confirms that* [GENERIC]"—emphasis added) appeared to convey the most powerful messages to readers.

In study 4b, when judging which sentences they would rather draw conclusions from, participants also judged bare generics to be less conclusive than framed generics [mean $= 3.22$, $t(102) = 5.85$, $P < 0.001$]. In contrast to the importance ratings, participants also judged bare generics to be slightly less conclusive than qualified and "some" nongenerics, $Ps < 0.05$. Participants appeared more confident about drawing conclusions when they were not just stipulated but were also said to have the backing of a research study.

General Discussion

The tendency to ignore variation, well-documented in participant recruitment, is echoed in scientific writing. In a sample of over 1,000 articles published in 11 psychology journals over a 2-y period, nearly 90% of articles included generics in the summary of research results (study 1). We found no evidence that this usage was warranted by stronger evidence, as it was uncorrelated with sample size. Instead, authors showed an overwhelming tendency to treat limited samples as supporting general conclusions, by means of universalizing statements (e.g., "Juvenile male offenders are deficient in emotion processing").

These generalizing statements covered a wide range of categories and constructs: People, women, children, adults, people with schizophrenia, self-promoters, early bilinguals, the brain, the human orbitofrontal cortex, statistical learning, mortality salience, parental warmth, social exclusion, zero-sum beliefs, emotion

regulation, effortful control, human decision-making, to name just a smattering. The present study may even underestimate the frequency of generics in scientific writing because we focused strictly on sentences describing study results in titles, highlights, and abstracts, thus excluding summaries of prior findings in the literature reviews, or implications in the discussion sections. On the other hand, because authors may resort to generics more often when word counts are restricted, the rates of generics may be higher in these more condensed portions of the paper. Common language practices thus contribute to a gap between the limitations of research evidence and the generality of conclusions.

These results—notable in their own right—have 2 further implications. First, generic language in scientific articles may lead readers to reach exaggerated conclusions. Past research found that generic sentences implied that a property was broadly true (28, 35, 45, 46) and conceptually central (37, 38), and that the category expressed was stable with inherent properties (41, 47). In the present work, studies 2 to 4 similarly revealed that both online workers and undergraduates studying psychology judged research summaries with generic language to be more important than nongeneric summaries, and under certain circumstances to be more generalizable and conclusive. At the same time, the present effects were small and subtle, and readers were more sensitive to language differences when multiple, converging cues were provided (e.g., "*Some* people with dysphoria *were* less sensitive to positive information in the environment" or "People with dysphoria *were* less sensitive to positive information, *in certain situations*"). Thus, in order to communicate potential limits on generality, authors may need to employ more explicit linguistic signals.

A second implication of study 1 is as a window on how scientists conceptualize data. Namely, the widespread use of generics suggests a widespread tendency on the part of authors to gloss over variation. Indeed, despite a near-universal tendency to report empirical results in generic terms, over 70% of the papers we sampled supplied no clear information about participants' race, SES, or language, consistent with other findings in the literature (48). Those articles providing no information about participants' SES were more likely to include generics in their results summaries (although papers that provided information about participants' language background were more likely to include generics than those that did not). Even when this information was provided, there was little consistency in how it was reported or even where it appeared. Writing as if variation does not exist downplays the importance of sampling broadly, and may lead to inappropriately aggregating across diverse groups or treating underrepresented groups as deficient (19, 49).

There likely are converging reasons for the ubiquity of generic language. Generics may often be an unwitting default, reflecting how generalizations are typically expressed in natural language. For example, when conducting this research, we were chastened to discover unintended generics in our own published writing (e.g., including claims about "children," rather than "a sample of English-speaking children raised in one middle-class, US community") (50). At the same time, bolder framing such as this may be a deliberate choice resulting from perverse incentives, when scientists have to convince journals, funding agencies, and the broader public of the importance of their research (14, 51, 52). In a climate in which submissions are routinely rejected by top journals and funding agencies, even a small effect of viewing generic summaries as more important could play a role in what is published or funded. It is also possible that an accumulation of small but consistent effects can build to result in larger disparities (53). Generic language may also result from writing guidelines and best practices that recommend using concise, compact language, and that provide generics as examples of "good writing" (54–56). Editorial policies requiring more condensed formats may also play a role. Recall that in study 1, shorter elements of an article (titles and highlights, compared to abstracts)

contained more generics and proportionately more unqualified generics than did longer elements of an article. Finally, there may be fashions that spread among a scientific community. For example, Rosner (51) found striking increases over time and differences across disciplines in the use of assertive sentence titles, defined as "complete sentences that assert a conclusion" and "[having] the form of an eternal truth" (which typically are generics; over 97% of the titles coded as generic in study 1 followed this format).

Although the present studies highlight the ubiquity of—and potential problems with—using generic language to describe research findings, it may be unrealistic to expect generics to disappear, given all of the factors reviewed above. A more fruitful approach may be that proposed by Simons, Shoda, and Lindsey (14), which they call "constraints on generality." The basic suggestion is that each published paper should include a statement that identifies and justifies the authors' beliefs about a study's target populations: The participants, stimuli, procedures, and cultural/historical contexts to which the results are likely to generalize (14). Being explicit about these assumptions reminds the reader of the limitations of the sample and provides a clear set of conditions that can be tested by others. The present studies have several such limitations. First, study 1 was restricted to psychology journals that included highlights or short summaries. It is an open question as to whether the same patterns would be obtained in other disciplines or for journals that do not require these additional brief elements. Second, most of our experimental studies did not include professionals in the field (e.g., students of psychology, research psychologists), who have greater knowledge and expertise and thus might more easily overlook how a paper was written in evaluating the importance or generality of the conclusions. (Note, however, that study 3c included

college students studying psychology, and their responses were comparable to those of the online sample.) Conversely, the participants in these studies—highly educated, English-speaking participants in the United States with internet access—may be more skilled at making subtle distinctions in language than other populations. Finally, participants only read one-sentence summaries of findings; thus, it is an open question whether readers might have different interpretations when presented with longer material with a mix of generic and nongeneric language, as is typically the case with published abstracts.

To conclude, further study is important to understand the scope and consequences that oversimplifying scientific findings has for the interpretation of those findings by experts, students, and the public. Psychology and many fields of scientific inquiry are confronting important questions as to the extent to which what was considered to be foundational knowledge in the field needs to be contextualized with attention to cultural differences (15, 19, 57), and the transparency and replicability of research efforts (6, 7). Considering the language used to communicate those findings may be one important step to raising awareness of these issues.

1. J. J. Arnett, The neglected 95%: Why American psychology needs to become less American. *Am. Psychol.* **63**, 602–614 (2008).
2. J. Henrich, S. J. Heine, A. Norenzayan, Most people are not WEIRD. *Nature* **466**, 29 (2010).
3. M. Nielsen, D. Haun, C. Kärtner, C. H. Legare, The persistent sampling bias in developmental psychology: A call to action. *J. Exp. Child Psychol.* **162**, 31–38 (2017).
4. S. F. Anderson, K. Kelley, S. E. Maxwell, Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychol. Sci.* **28**, 1547–1562 (2017).
5. K. S. Button et al., Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
6. M. R. Munafò et al., A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
7. Open Science Collaboration, PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science* **349**, aac4716 (2015).
8. Y. Weinstein, M. A. Sumeracki, Are twitter and blogs important tools for the modern psychological scientist? *Perspect. Psychol. Sci.* **12**, 1171–1175 (2017).
9. P. Sumner et al., The association between exaggeration in health related science news and academic press releases: Retrospective observational study. *BMJ* **349**, g7015 (2014).
10. S. Oyama, *The Ontogeny of Information: Developmental Systems and Evolution* (Duke University Press, Durham, NC, 2000).
11. S. Oyama, "The nurturing of natures" in *On Human Nature: Anthropological, Biological, and Philosophical Foundations*, A. Grunwald, A. Gutmann, E. M. Neumann-Held, Eds. (Springer, Berlin, 2002), pp. 163–170.
12. R. S. Siegler, *Emerging Minds: The Process of Change in Children's Thinking* (Oxford University Press, New York, 1996).
13. L. F. Barrett, Categories and their role in the science of emotion. *Psychol. Inq.* **28**, 20–26 (2017).
14. D. J. Simons, Y. Shoda, D. S. Lindsay, Constraints on generality (COG): A proposed addition to all empirical papers. *Perspect. Psychol. Sci.* **12**, 1123–1128 (2017).
15. B. Rogoff, *The Cultural Nature of Human Development* (Oxford University Press, New York, 2003).
16. K. D. Gutiérrez, B. Rogoff, Cultural ways of learning: Individual traits or repertoires of practice. *Educ. Res.* **32**, 19–25 (2003).
17. N. Akhtar, V. K. Jaswal, Deficit or difference? Interpreting diverse developmental paths: An introduction to the special section. *Dev. Psychol.* **49**, 1–3 (2013).
18. M. Callanan, S. Waxman, Commentary on special section: Deficit or difference? Interpreting diverse developmental paths. *Dev. Psychol.* **49**, 80–83 (2013).
19. D. Medin, W. Bennis, M. Chandler, Culture and the home-field disadvantage. *Perspect. Psychol. Sci.* **5**, 708–713 (2010).
20. G. N. Carlson, F. J. Pelletier, Eds., *The Generic Book* (University of Chicago Press, 1995).
21. O. Dahl, "12 The marking of the episodic/generic distinction in tense-aspect systems" in *The Generic Book*, G. N. Carlson, F. J. Pelletier, Eds. (University of Chicago Press, Chicago, IL, 1995), pp. 412–426.
22. S. A. Gelman, *The Essential Child: Origins of Essentialism in Everyday Thought* (Oxford University Press, New York, 2003).
23. J. Lawler, "Tracking the generic toad" in *Proceedings from the Ninth Regional Meeting of the Chicago Linguistic Society*, (Chicago Linguistic Society, Chicago, 1973) pp. 320–331.
24. S.-J. Leslie, Generics and the structure of the mind. *Philos. Perspect.* **21**, 375–403 (2007).
25. S.-J. Leslie, Generics: Cognition and acquisition. *Philos. Rev.* **117**, 1–47 (2008).
26. I. Prasada, Acquiring generic knowledge. *Trends Cogn. Sci. (Regul. Ed.)* **4**, 66–72 (2000).
27. S. A. Gelman, S. O. Roberts, How language shapes the cultural inheritance of categories. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7900–7907 (2017).
28. S. A. Graham, S. A. Gelman, J. Clarke, Generics license 30-month-olds' inferences about the atypical properties of novel kinds. *Dev. Psychol.* **52**, 1353–1362 (2016).
29. M. A. Hollander, S. A. Gelman, J. Star, Children's interpretation of generic noun phrases. *Dev. Psychol.* **38**, 883–894 (2002).
30. B. Mannheim, S. A. Gelman, C. Escalante, M. Huayhua, R. Puma, A developmental analysis of generic nouns in Southern Peruvian Quechua. *Lang. Learn. Dev.* **7**, 1–23 (2010).
31. T. Tardif, S. A. Gelman, X. Fu, L. Zhu, Acquisition of generic noun phrases in Chinese: Learning about lions without an '-s'. *J. Child Lang.* **39**, 130–161 (2012).
32. S.-J. Leslie, S. A. Gelman, Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognit. Psychol.* **64**, 186–214 (2012).
33. R. P. Abelson, D. E. Kanouse, "Subjective acceptance of verbal generalizations" in *Cognitive Consistency*, F. Feldman, Ed. (Academic Press, New York, 1966), pp. 171–197.
34. A. C. Brandone, A. Cimpian, S.-J. Leslie, S. A. Gelman, Do lions have manes? For children, generics are about kinds rather than quantities. *Child Dev.* **83**, 423–433 (2012).
35. A. Cimpian, A. C. Brandone, S. A. Gelman, Generic statements require little evidence for acceptance but have powerful implications. *Cogn. Sci.* **34**, 1452–1482 (2010).
36. R. R. Schmeck, D. Lockhart, Introverts and extraverts require different learning environments. *Educ. Leadersh.* **40**, 54–55 (1983).
37. A. Cimpian, E. M. Markman, Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition* **113**, 14–25 (2009).
38. M. A. Hollander, S. A. Gelman, L. Raman, Generic language and judgements about category membership: Can generics highlight properties as central? *Lang. Cogn. Process.* **24**, 481–505 (2009).

39. S. Prasada, E. M. Dillingham, Principled and statistical connections in common sense conception. *Cognition* **99**, 73–112 (2006).

40. S. Prasada, E. M. Dillingham, Representation of principled connections: A window onto the formal aspect of common sense conception. *Cogn. Sci.* **33**, 401–448 (2009).

41. M. Rhodes, S.-J. Leslie, C. M. Tworek, Cultural transmission of social essentialism. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13526–13531 (2012).

42. A. Orvell, E. Kross, S. A. Gelman, How "you" makes meaning. *Science* **355**, 1299–1302 (2017).

43. D. Wodak, S.-J. Leslie, M. Rhodes, What a loaded generalization: Generics and social cognition. *Philos. Compass* **10**, 625–635 (2015).

44. E. Pain, How to (seriously) read a scientific paper. *Science*, 10.1126/science.caredit. a1600047. (2016).

45. S. A. Gelman, J. R. Star, J. Flukes, Children's use of generics in inductive inferences. *J. Cogn. Dev.* **3**, 179–199 (2002).

46. S. A. Graham, S. L. Nayer, S. A. Gelman, Two-year-olds use the generic/nongeneric distinction to guide their inferences about novel kinds. *Child Dev.* **82**, 493–507 (2011).

47. S. A. Gelman, E. A. Ware, F. Kleinberg, Effects of generic language on category content and structure. *Cognit. Psychol.* **61**, 273–301 (2010).

48. M. S. Rad, A. J. Martingano, J. Ginges, Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11401–11405 (2018).

49. D. Medin, B. Ojalehto, A. Marin, M. Bang, Systems of (non-)diversity. *Nat. Hum. Behav.* **1**, 0088 (2017).

50. S. A. Gelman, Psychological essentialism in children. *Trends Cogn. Sci. (Regul. Ed.)* **8**, 404–409 (2004).

51. J. L. Rosner, Reflections of science as a product. *Nature* **345**, 108 (1990).

52. N. Di Girolamo, R. M. Reynders, Health care articles with simple and declarative titles were more likely to be in the Altmetric Top 100. *J. Clin. Epidemiol.* **85**, 32–36 (2017).

53. V. Valian, *Why So Slow?: The Advancement of Women* (MIT Press, Cambridge, MA, 1999).

54. R. V. Kail, *Scientific Writing for Psychology: Lessons in Clarity and Style* (SAGE Publications, Thousand Oaks, CA, 2014).

55. H. L. Roediger, III, Twelve tips for authors. *APS Obs.* 20, (2007). https://www.psychologicalscience.org/observer/twelve-tips-for-reviewers, Accessed 1 July 2018.

56. C. J. Weinberger, J. A. Evans, S. Allesina, Ten simple (empirical) rules for writing science. *PLOS Comput. Biol.* **11**, e1004205 (2015).

57. G. Solis, M. Callanan, Evidence against deficit accounts: Conversations about science in Mexican heritage families living in the United States. *Mind Cult. Act.* **23**, 212–224 (2016).

PSYCHOLOGICAL AND COGNITIVE SCIENCES