Constraints on Generality (COG): A Proposed Addition to All Empirical Papers

Daniel J. Simons[1], Yuichi Shoda[2], and D. Stephen Lindsay[3]
[1]University of Illinois, [2]University of Washington, and [3]University of Victoria

Abstract
Psychological scientists draw inferences about populations based on samples—of people, situations, and stimuli—from those populations. Yet, few papers identify their target populations, and even fewer justify how or why the tested samples are representative of broader populations. A cumulative science depends on accurately characterizing the generality of findings, but current publishing standards do not require authors to constrain their inferences, leaving readers to assume the broadest possible generalizations. We propose that the discussion section of all primary research articles specify Constraints on Generality (i.e., a "COG" statement) that identify and justify target populations for the reported findings. Explicitly defining the target populations will help other researchers to sample from the same populations when conducting a direct replication, and it could encourage follow-up studies that test the boundary conditions of the original finding. Universal adoption of COG statements would change publishing incentives to favor a more cumulative science.

If you are like most psychologists, you hope to draw conclusions from your research that go beyond the particulars of your studies. Your participants represent a population of people. Your tasks assess underlying constructs broader than the tasks themselves. Your materials draw from a class of similar materials that you could have used. Your procedure is one of many ways you could have addressed the same scientific questions. Scientific inference is about drawing general conclusions from specific situations. Nearly all research in psychology draws conclusions about a population of people, behaviors, and contexts from a sample of participants, materials, and procedures. You know this. Everyone knows this. It goes without saying. And that is a problem.

Psychologists rarely identify the target populations for their inferences, even when they treat their participants and materials as random effects in their analyses. Even fewer justify their (often implicit) claims of generality. For example, psychology articles often report findings as if they apply to all of humanity even if the tested sample was notably WEIRD (Western, Educated, Industrialized, Rich, and Democratic; Henrich, Heine, & Norenzayan, 2010). If pressed, though, most researchers likely would acknowledge that they lack evidence for unconstrained generality.

This failure to specify the target populations—or to state the degree of uncertainty about how broadly the reported findings will generalize—contributes to myriad problems in scientific communication, including tensions when a "failed" replication differs in subtle ways from an original study. When an article does not identify the target populations, other researchers may reasonably assume that the finding applies broadly and that any study that samples from those broadest populations provides a test of the same effect. In contrast, when a paper does identify the target populations and specifies constraints on the generality of the findings, researchers conducting direct replications will deliberately sample from the target populations, leading to a more appropriate test of the reliability of the original claims. Scientific knowledge accumulates as we learn how broadly our findings apply and how they are limited.

Corresponding Author:
Daniel Simons, Department of Psychology, University of Illinois, 603 E. Daniel Street, Champaign, IL 61820
E-mail: dsimons@illinois.edu

Our proposal: The discussion section of all articles describing empirical research should include a statement of the *Constraints on Generality* (a "COG" statement) that explicitly identifies and justifies the target populations for the reported findings. The gears of science will turn more effectively with functional COGs, and we call on editors and reviewers to request a COG statement when one is lacking.

The idea of specifying constraints on generality is not new. We and others have argued for better specification of the target population for generalizability and for clearer identification of the parameters critical for observing an effect (e.g., Brandt et al., 2014; Clark, 1973; Greenwald, Pratkanis, Leippe, & Baumgardner, 1986; Henrich et al., 2010; Kukull & Ganguli, 2012; Shoda, Wilson, Whitsett, Lee-Dussud, & Zayas, 2014; Simons, 2014; Wells & Windschitl, 1999). Formalizing these goals and incorporating them as a standard part of empirical papers will help other scientists, students, journalists, and the public understand a study's results and implications.

Method sections of empirical papers already describe the sample and procedures used in a study, but those details differ from the contents of a COG statement. A method section describes the proximal populations—the immediate source of the sample for that study. For example, a method section might note that participants were recruited from the psychology subject pool at the University of Illinois. However, most psychology researchers presumably are not interested in generalizing the results of their studies to the subject pool at their university. Instead, they believe, based on empirical evidence or theoretical considerations, that their proximal populations are representative of broader groups, such as all psychology undergraduates, all undergraduates in the United States, all adults, or even all mammals. A COG statement specifies your intended target population and the basis for believing that your sample is representative of it; it justifies why the subjects, materials, and procedures described in the method section are representative of broader populations.

There are (at least) two grounded reasons for expecting generality or specificity: (a) direct empirical evidence and (b) theoretical predictions derived from past research. Theoretical claims of generality—those not based on compelling empirical evidence—should be subject to further empirical examination including replications; they should be treated as "not yet supported."

Explicit statements about generality will increase the value of both direct and conceptual replications. Direct replications will be more likely to draw from the same target populations as the original study and correctly reproduce the critical test of the effect. Conceptual replications can use the information specified in a COG statement to vary the sampled populations, thereby testing the generality of an effect. The presence of a COG statement should facilitate publishing of such follow-up research because authors can document the need for that research whether or not it produces the same findings. Over the long term, broad adoption of COG statements should lead to a more cumulative understanding of the scope of the effects we study.

Incentives for Appropriate COG Statements

The current publishing model incentivizes authors to make the strongest possible claims of generality; broadly generalizable findings are more likely to be published and more likely to be influential. Why then should individual authors constrain their claims? And why should reviewers and editors favor an appropriately constrained claim over a broad, "newsworthy" one?

First, COG statements decrease the chance that your claims will prove to be embarrassingly more limited than you had implied; it could be seen as misleading if a paper implied generality to all of humanity when its findings plausibly might generalize only to college sophomores at a public university in Illinois. With the increasing prevalence of postpublication review, curation of related follow-up research (e.g., curatescience.org, psychological filedrawer.org), social media discussion of findings, and a growing emphasis on replication, the credibility-damaging consequences of publishing papers that turn out to have wildly exaggerated claims of generality have increased. By specifying the target population of participants, stimuli/materials, and procedures, and noting uncertainty where appropriate, a COG statement protects authors from inadvertently making overly broad claims.

Second, a COG statement increases the likelihood that other researchers will succeed when replicating your findings; they can accurately sample from your intended target population. Moreover, if you note your uncertainty about the generality of your results, those replicating your research will acknowledge the ways in which your original claims were tentative. In effect, the COG statement counts as a preregistered commentary on future replication studies.

Third, a COG statement may inspire follow-up studies that build upon your results by testing their generality in populations you did not test. Then a failure to observe the effect in samples from different populations will be seen as helping to establish boundaries on generality rather than as a "replication failure." COG statements also encourage reviewers and editors to be more receptive to "next-step" studies that test the constraints you identified; authors (including you) can point to the COG statement to motivate and justify a new study.

Editors have different incentives than individual authors: They want to publish influential work that will be highly cited, but they also want to publish robust, reliable research that will advance the field as a whole. They are the gatekeepers for our field's credibility, and they can adopt a broader, long-view perspective on the state of the field. Together with academic societies and publishers, editors can work to change the incentive structure in the field. By requiring COG statements for a journal, editors can ensure that all papers will be evaluated on even footing and that honesty and accuracy in statements of generality are valued.

Once COG statements are included in all papers, editors will favor manuscripts with a well-justified COG statement that explicitly grounds its claims of generality. If a paper can justify broad generalization, then all incentives will favor publication. If it cannot, then editors will evaluate whether the claims it can make are sufficiently important to justify publication. If COG statements incentivize cumulative follow-up research, they will fulfill some of the other goals of editors as well: increased citations, greater reliability, greater influence.

Reviewers have a more limited set of goals than editors. Their primary objective is quality control: Is the research reported in a manuscript rigorous and are the conclusions sufficiently supported to merit publication? Although reviewers may favor papers that are interesting and provocative, they also value rigor, honesty, and accuracy. COG statements help them to verify that the paper's conclusions are justified by the evidence and theory. They might find a paper with a highly restrictive COG statement to be less groundbreaking and, as a result, less interesting. But they also will recognize that it likely makes a more accurate and verifiable contribution to the literature than one making unfettered and unjustified generalizations. A COG statement helps reviewers make an informed recommendation.

In the absence of a COG statement, readers may take the implied claims of generality and practical relevance at face value; they may go unchecked. If our science were more cumulative and self-correcting, implying broad generalization might be justifiable; strong claims would motivate new research, which would then constrain those claims. That "assume general, constrain later" approach has been a failure: Strong, incorrect claims have a way of persisting even when later research reveals their limits (e.g., Bangerter & Heath, 2004; for discussion, see Lilienfeld, Marshall, Todd, & Shane, 2014), and replication attempts that vary some aspect of the procedure and fail to observe the same result are often taken not as constraints on the generality of the original findings but as evidence that the effect does not exist (Pashler & Harris, 2012).

A COG statement makes claims about scope explicit and easier to evaluate. Empirically grounded claims in a COG statement are verifiable: Reviewers and editors can ask whether the cited evidence justifies the claim. If a manuscript claims generality to all of humanity but cites no evidence that the methods are likely to apply that broadly, reviewers can (and should) require authors to support their assertions or to limit them. Similarly, theoretically grounded claims of generality that lack direct empirical evidence can be evaluated just like any other theoretical statement. That is, reviewers and editors can assess the plausibility and credibility of the claimed generality.

What Should Be Included in a COG statement

A COG statement defines the scope of the conclusions that are justified by your data. It clarifies which aspects of your sample of participants, materials, and procedures should be preserved in a direct replication and identifies both those aspects believed to be crucial to observing the effect and those thought to be irrelevant. Of course, we cannot know in advance all of the factors that might moderate an effect, so a COG statement cannot be an exhaustive list. We suggest the following principles in deciding whether or not to mention a factor.

1. Known empirical or theoretical boundary conditions should be included except those that would be obvious to almost all readers based on common knowledge or common sense (e.g., there would be no need to specify the ability to hear as a boundary condition for a study of auditory discrimination).
2. Boundary conditions that are tied to the substance of the study should be mentioned even if their constraining role lacks direct empirical or theoretical support. For example, a study about political party affiliation conducted immediately before an election should indicate that the results might depend on that political climate.
3. Factors that experts in a discipline might consider to be important (i.e., known unknowns) should be noted, even if their role in constraining generality is untested. For example, for a study manipulating gender markers in language, it would be reasonable to expect that the results would not generalize to speakers of languages that lack gender markers, even if the same conceptual questions would be meaningful in those languages.
4. Other factors need not be listed. To make it explicit that a COG statement describes the known or anticipated limits on the finding and not possible mediation by "unknown unknowns," we recommend that authors include a 19-word boilerplate sentence at the end of the COG statement: "We

have no reason to believe that the results depend on other characteristics of the participants, materials, or context." Future research might well uncover such dependencies, and researchers should seek them because doing so refines our understanding of the proposed mechanisms, turning the gears of science.

A COG statement should take these principles into account when specifying the target populations and constraints for each of the following aspects of their study.

- *Participants*. Discuss how your proximal sample of participants is representative of a broader target population. If you tested undergraduates, do you believe the findings apply only to students at your university? To students at comparable universities? To students anywhere? To all adults? To all mammals? What would another researcher need to do to verify that their participants were drawn from the same target population as yours? If your COG statement specifies a target population of all adults, you are accepting that a replication with any sample of adults would constitute a direct test of the same effect.
- *Materials/Stimuli*. Define the class of materials/stimuli to which your finding should generalize. What are the critical features of your materials that must be maintained to measure the same construct? What measurements are necessary to verify that any new materials tap the same target population as your materials?
- *Procedures*. What aspects of your procedures must be followed closely to reproduce the effect? What broader class of procedures should produce the same results? For example, would future studies need to test participants in your lab? Would researchers need to use the same computers, and if so, have you provided enough detail in your method section for them to do so? Will the effect work only if participants are tested in isolated cubicles, or will it also work in a large classroom setting or in a shopping mall? Can any undergraduate administer the tasks or does the study require special training? What checks are needed to ensure that the procedures match the broader population of procedures that can produce the effect?
- *Historical/temporal specificity*. Does the effect depend on cultural norms that might change over time? For example, results of studies involving attitudes about same-sex marriage in a study conducted in the 1990s might differ from those in a study conducted in 2017. Similarly, studies of attitudes about politics might differ when measured right before or after an election. What aspects of

the temporal or historical context need to be stable to observe the effect? Can you anticipate and specify any differences in the historical or temporal context that might affect whether or not other researchers would observe the same effect?

Sample Cases

In this section, we provide example COG statements for three of our own findings. Given that COG statements justify the link between the proximal populations of a study and the target populations for generalization, they tend to vary substantially across studies. Just as the method sections of highly similar studies overlap in content, COG statements for closely linked studies might adopt similar language. But, as these examples show, the contents are more varied and substantive than empty boilerplate statements such as "more research is needed."

Example 1

Simons (2013) measured overconfidence and memory in bridge players. Participants predicted their performance relative to other pairs of players prior to the start of each duplicate bridge session at a bridge club. At the end of the study, they estimated how well they had done, on average, across all of the bridge sessions. On average, the players expected to perform better than they actually did. Yet, they accurately remembered their average scores for the sessions they had played. In other words, they knew how well they actually performed, but they were overly optimistic about how they would perform in each session. This is how we would write a COG statement for this article (adapting prose from a statement of limitations, p. 603).

> Our finding provides evidence of the Dunning-Kruger effect (Kruger & Dunning, 1999) in participants who are aware of their relative skill. Given that this "better than average" effect has been observed for a diverse range of participants in a wide range of tasks (including unpublished evidence from our own laboratory with chess players), we expect our result with bridge players to generalize to other domains in which players regularly compete against the same group of players in games of skill. However, given that relative performance in any given session of duplicate bridge involves some luck, the pattern of results—optimistic predictions but accurate memory—might hold only for games that involve both skill and luck. A direct replication would test bridge players in sessions that include players with skill levels ranging from relative novice to expert in the context of their regular bridge game (i.e., the players should play with and against each other at least weekly and should be familiar with

the skill level of the other players in each session). Participants should be blind to the predictions made by other players to avoid having knowledge of those predictions affect their play. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Example 2

Using the "Highly Repeated Within Person" paradigm, Whitsett and Shoda (2014) examined, for each participant, the relationship between support seekers' expression of distress and participants' willingness to support them. On average, the greater a support seeker's expression of distress, the greater the participants' willingness to support them. But for a sizable minority of participants, statistically significant effects in the opposite direction were observed: Their willingness to help reliably decreased rather than increased in response to greater distress expressed by support seekers. This is how we would write a COG statement for this finding.

> The stimuli consisted of a large number of video clips in which a large number of different undergraduates sampled from the subject pool at the University of Washington each expressed mild distress in their own way. Thus, we expect the results to generalize to situations in which participants view similar video clips, as long as manipulation checks indicate the clips depict a variety of ways in which people express mild distress. Unpublished studies from our laboratory resulted in similar results despite variations in the testing context (e.g., different research assistants). Consequently, we do not expect such variations to matter. We believe the results will be reproducible with students from similar subject pools serving as participants. However, we do not have evidence that the findings will occur outside of laboratory settings. The distress expressed in the video clips was triggered by a specific laboratory induction, and we lack evidence showing that the results will generalize to expressions of distress in response to other situations. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Example 3

Lindsay, Hagen, Read, Wade, and Garry (2004) reported a study in which undergraduates were exposed to suggestions that, when they were in Grade 1 or 2, they and an accomplice put Slime in their teacher's desk. By the end of a three-phase procedure that unfolded over a week or so, a substantial percentage of subjects were categorized by judges as "remembering" this event, especially so for subjects who had been given a copy of their class photograph as a memory cue. The article discussed limitations on the generalizability of the results, but not in a single COG statement. It made no mention of the potential role of the accomplice in moderating the photo effect, nor of the potential role of the skills of the interviewer in producing the high rate of apparent false memories. Here is the COG statement we would write for that article today.

> The results from our no-photo condition converge with prior evidence that combining a plausible narrative attributed to a family member with social pressure, demand characteristics, and sustained memory recovery techniques can lead a substantial percentage of undergraduate subjects to appear to remember a childhood pseudoevent. The relative contributions of these components is unclear. Moreover, the likelihood of false memory reports is affected by numerous variables including the nature of the suggested event (see Lindsay & Read, 2001); the absolute rate of false memories in our study should not be used to predict the probability of false memories of childhood sexual abuse. Moreover, the very high false memory rate in our photo condition may be specific to this suggested event and photo. Our suggested event involved an accomplice, and we speculate that this may have amplified the photo effect by helping subjects imagine the event. We do expect, though, that the rates of false memory for similar types of events (i.e., events with a similar rate of false memory) should generally be higher with a photo memory prompt than without one provided that the photo supplies information that participants can use to imagine the suggested event. It must be noted, however, that our sample sizes were modest, especially given the nature of the measures and the design, so the absolute rates of false memories that we observed might well differ in replications on statistical grounds. We speculate that asking subjects about increasingly remote events (a Grade 5 or 6 event and then a Grade 3 or 4 event before asking about the Grade 1 or 2 pseudoevent) may also have increased false memory rates. Finally, all subjects were tested by the second author, who was (in the judgment of the first author) adept at presenting the suggestions in a compelling way and motivating the subjects (who were younger than she) to work hard at remembering the pseudoevent. We speculate that these skills increase the likelihood of false memory reports. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

Conclusion

By identifying and justifying your beliefs about the target populations for your study—not just of participants, but also of stimuli, procedures, and cultural/historical contexts—a COG statement helps other scientists evaluate the scope of your claims and it motivates future research testing the boundary conditions of your findings. It ensures that direct replication attempts will sample from your intended target population, and it defines where additional research on the generality of an effect is needed. Nothing about a COG statement is inherently novel—it makes explicit and verifiable what has typically been unstated and unjustified. Writing a COG statement prompts authors to consider and articulate their target populations, and reading a COG statement prompts other researchers to evaluate which claims of generality already have empirical support and which do not.

The most exciting psychological research changes the way psychologists think about important questions; the findings are both surprising and real (Fiedler, 2017). We see our proposal for COG statements as part of a larger movement to nudge scientific psychology away from publishing splashy, one-off claims of dubious generality and toward an emphasis on rigor and understanding of boundary conditions. COG statements are not a panacea, but they can turn the gears of the scientific engine and advance the field toward more robust, cumulative evidence.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

Bangerter, A., & Heath, C. (2004). The Mozart effect: Tracking the evolution of a scientific legend. *British Journal of Social Psychology*, *43*, 605–623.

Brandt, M. J., Ijzerman, H., Dijksterhuis, A., Farach, F. J., Geller, J., Giner-Sorolla, R., & Van't Veer, A. (2014). The replication recipe: What makes for a convincing replication? *Journal of Experimental Social Psychology*, *50*, 217–224.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science*, *12*, 46–61.

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, *93*, 216–229.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*, 29.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121–1134.

Kukull, W. A., & Ganguli, M. (2012). Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology*, *78*, 1886–1891.

Lilienfeld, S. O., Marshall, J., Todd, J. T., & Shane, H. C. (2014). The persistence of fad interventions in the face of negative scientific evidence: Facilitated communication for autism as a case example. *Evidence-Based Communication Assessment and Intervention*, *8*(2), 62–101.

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science*, *15*, 149–154.

Lindsay, D. S., & Read, J. D. (2001). The recovered memories controversy: Where do we go from here? In G. Davies & T. Dalgleish (Eds.), *Recovered memories: Seeking the middle ground* (pp. 71–94). London, England: Wiley.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*, 531–536.

Shoda, Y., Wilson, N. L., Whitsett, D. D., Lee-Dussud, J., & Zayas, V. (2014). The person as a Cognitive-Affective Processing System: From quantitative idiography to cumulative science. In M. L. Cooper & R. J. Larsen (Eds.), *Handbook of personality processes and individual differences* (pp. 491–513). Washington, DC: APA Press.

Simons, D. J. (2013). Unskilled and optimistic: Overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin & Review*, *20*, 601–607. doi:10.3758/s13423-013-0379-2

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*, 76–80.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115–1125.

Whitsett, D. D., & Shoda, Y. (2014). Examining the heterogeneity of the effects of situations across individuals does not require a priori identification and measurement of individual difference variables. *Journal of Experimental Social Psychology*, *50*, 94–104.