

## COMMENT

# Effects of Categorical Labels on Similarity Judgments: A Critical Evaluation of a Critical Analysis: Comment on Noles and Gelman (2012)

Vladimir M. Sloutsky  
The Ohio State University

Anna V. Fisher  
Carnegie Mellon University

Noles and Gelman (2012) attempt to critically reevaluate the claim that linguistic labels affect children's judgments of visual similarity. They report results of an experiment that used a modified version of Sloutsky and Fisher's (2004) task and conclude that "labels do not generally affect children's perceptual similarity judgments; rather, children's reliance on labels to make similarity judgments appears to be attributable to flaws in the methodological approaches used in prior studies" (p. 890). In this comment we demonstrate that these conclusions are based on inadequate analyses and reporting of statistics and are inconsistent with Noles and Gelman's own data. Instead, their data show a perfect replication of Sloutsky and Fisher (2004) and present further evidence that linguistic labels affect similarity judgment of young children.

*Keywords:* categorization, similarity, language, cognitive development

In their brief report, Noles and Gelman (2012) attempt to critically reevaluate the claim that linguistic labels affect children's judgments of visual similarity. This claim is based on the finding that children (but not adults) tend to consider items as looking more similar when the objects are accompanied by the same label than when these objects are accompanied by different labels or when labels are not provided (e.g., Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999). Accordingly, Noles and Gelman replicate findings reported by Sloutsky and Fisher (2004, Experiment 1) under somewhat different experimental conditions.

Why are this claim and its reevaluation important? Whereas there is a general agreement that linguistic labels affect inductive generalization, there is little agreement about the underlying mechanism. Do labels function as features contributing to the overall perceived similarity or as category markers representing category membership? This issue is particularly contentious with respect to (a) the role of linguistic labels early in development and (b) developmental changes in the role of labels in categorization and induction.

It has been argued that from early in development children expect that labels communicate category membership of objects

(e.g., Gelman, 2003; Gelman & Coley, 1990; Waxman & Markow, 1995). Alternatively, it has been suggested that early in development, labels are part of input rather than category markers—they are akin to other perceptual features of objects (Deng & Sloutsky, 2012; Sloutsky, 2003, 2010; Sloutsky & Fisher, 2004, 2012; Sloutsky & Lo, 1999; Sloutsky, Lo, & Fisher, 2001).

Whether labels contribute to similarity judgment is important because this finding helps distinguishing between the two theoretical possibilities: If labels are category markers, they should not contribute to perceptual similarity of items, whereas if they are part of input, they may. Therefore, findings that labels contribute to similarity judgments of young children present a challenge to the label-as-category-marker account. Of course, these are not the only problematic findings for this account (see Sloutsky & Fisher, 2005, for an overview, as well as more recent work: Deng & Sloutsky, 2012; Fisher, 2010; Fisher, Matlen, & Godwin, 2011; Sloutsky & Fisher, 2012).

In their study, Noles and Gelman (2012) focused on effects of labels on similarity judgment and concluded that "labels do not generally affect children's perceptual similarity judgments" (p. 890). Although we think that Noles and Gelman's attempt to replicate and critically evaluate previous findings is important, our examination of the reported findings and our own reanalysis of the Noles and Gelman data suggest that the conclusions presented in the article are not supported by the data. In what follows, we present a brief summary of Noles and Gelman's findings and explain the disconnect between these findings and conclusions. We also provide results of additional analyses suggesting that Noles and Gelman perfectly replicated effects of labels on young children's similarity judgment reported by Sloutsky and Fisher (2004; see also Sloutsky & Lo, 1999).

Recall that Sloutsky and Fisher (2004) demonstrated that 4–5-year-olds (but not adults) judged two items to look more alike when the items were accompanied by the same label than when

---

Vladimir M. Sloutsky, Department of Psychology and Center for Cognitive Science, The Ohio State University; Anna V. Fisher, Department of Psychology, Carnegie Mellon University.

We thank Nick Noles and Susan Gelman for sharing their data with us, and members of the Cognitive Development Lab at Ohio State for discussing this comment with Vladimir M. Sloutsky. Writing of this comment was supported by National Science Foundation Grant BCS-0720135 and by National Institutes of Health Grant R01HD056105 to Vladimir M. Sloutsky.

Correspondence concerning this article should be addressed to Vladimir M. Sloutsky, Department of Psychology and Center for Cognitive Science, 208C Ohio Stadium East, 1961 Tuttle Park Place, The Ohio State University, Columbus, OH 43210. E-mail: sloutsky.1@osu.edu

labels were not presented. First, Noles and Gelman (2012) hypothesized that children may have been confused by the wording of the similarity judgment question used by Sloutsky and Fisher. Note that Sloutsky and Fisher conducted a control experiment to eliminate this possibility (Experiment 1A); nevertheless, it is useful to have an additional (and more stringent) control. To control for possible confusion effects, Noles and Gelman included a label-plus-training condition in their design. In this condition, children were provided with 12 training trials clarifying that the similarity question referred to visual similarity. There were two additional conditions in the design: label-minus-training (in which the similarity question format closely followed the one used by Sloutsky and Fisher) and no-label (in which no labels were provided). In all three conditions, children were asked to perform similarity judgment with a subset of visual stimuli used by Sloutsky and Fisher.

Next Noles and Gelman (2012) reduced the number of levels of similarity compared with Sloutsky and Fisher. Note that to be able to capture effects of labels quantitatively, Sloutsky and Fisher (2004) used four levels of similarity ratios (SR), whereas Noles and Gelman selected two of those (i.e., SR = 1 and SR = 9). In the former case, both test items were equally similar to the target (Noles and Gelman called these *identical triads*), whereas in the latter case, one test item was substantially more similar to the target than was the other test item (Noles and Gelman called these *discriminable triads*). Finally, Noles and Gelman examined effects of labels on similarity judgment in two different orders of presentation. In Order 1, children were first presented with SR = 9 triads (in which label match was pitted against perceptual similarity) and then with SR = 1 triads (in which matching labels did not compete against perceptual similarity). In Order 2, this sequence was reversed. Note that Sloutsky and Fisher randomized the order of presentation of triads from the four levels of SR.

Noles and Gelman (2012) found the following:

1. An effect of similarity ( $F = 131.78$ ,  $\eta_p^2 = .622$ ); similarity judgment is a function of similarity ratio;
2. An effect of condition ( $F = 6.61$ ,  $\eta_p^2 = .142$ ); post hoc Bonferroni tests suggested that label-plus-training and label-minus-training conditions were statistically equivalent ( $p = .86$ ), whereas both differed significantly from the no-label condition ( $p < .05$ ); therefore, this effect reflects influence of matching labels on perceived similarity;
3. An effect of order ( $F = 8.76$ ,  $\eta_p^2 = .099$ ); items were judged more similar in Order 2 than in Order 1;
4. Order by similarity of interaction ( $F = 12.89$ ,  $\eta_p^2 = .139$ ); similarity ratio effect was moderated by order of presentation.

At the same time, neither the order-by-condition interaction ( $F = 1$ ,  $\eta_p^2 = .024$ ) nor the three-way interaction ( $F = 0.371$ ,  $\eta_p^2 = .009$ ) approached significance. Incidentally, the former is not reported in the article; this result and associated statistics transpired in our own reanalysis of the Noles and Gelman (2012) data.

Given the nonexistent order-by-condition interaction, one cannot conclude that effects of labels are greater in one order than in

the other (because this implies an interaction). Statistical textbooks treat this issue in a straightforward way: If an omnibus test is not significant, a posteriori tests should not be made (Kirk, 1995, p. 176). And if a posteriori tests are made, they should be treated as exploratory at best. Yet, Noles and Gelman conclude that labels influence perceptual similarity only under the “adverse” order of conditions (i.e., Order 2).

While ignoring the nonexistent order-by-labeling interaction and the three-way interaction, Noles and Gelman (2012) weighed heavily a separate a posteriori analysis within the “nonadverse” Order 1. Unfortunately, although conducting these exploratory analyses, Noles and Gelman do not report effects of labeling condition within Order 1. Our analysis of Noles and Gelman’s data revealed a respectable effect size for labeling in Order 1 ( $\eta_p^2 = .102$ ), which is bigger or equivalent to that of the reported order effect ( $\eta_p^2 = .099$ ). However, it stands to reason that if the effect size of .099 is large enough for the order effect (see p. 894 for references to the order effect), then the effect size of .102 should be large enough for the labeling effect in Order 1. All Noles and Gelman had to do to see that was to collapse across the two statistically equivalent conditions (i.e., label-plus-training and label-minus-training), thus increasing power, without changing the effect size. If one does that, it becomes clear that the effect of labeling is significant even in the nonadverse Order 1 ( $F = 4.51$ ,  $p < .05$ ). In addition, to see that their conclusions were unwarranted, Noles and Gelman could have done more extensive exploratory analyses by focusing on what should be a critical contrast: a comparison of the label-plus-training condition (which clarifies that the target question refers to visual similarity) and the no-label condition within Order 1: These analyses also point to a sizable effect of labeling within Order 1 ( $\eta_p^2 > .132$ ). Or they could have computed effect size estimates (i.e., Cohen’s  $d$ ) for all mean comparisons of the label versus no-label conditions. These estimates are presented in Table 1. Data in the table point to medium-to-large effect sizes for every condition included in Noles and Gelman’s design: For each of the eight comparisons, there was a nontrivial effect in the direction predicted by Sloutsky and Fisher (2004).

We therefore ask the following: Given (a) the omnibus statistics indicating that the labeling condition effect is significant but the order-by-condition interaction or a three-way interaction are not, (b) sizable effect of labeling condition for each order, and (c) respectable effects sizes for labeling effect in every condition (see

Table 1  
Cohen’s  $d$ s for Label Effects (i.e., Pairwise Comparisons for Label and No-Label Conditions) for Each Order and Similarity Ratio in the Noles and Gelman (2012) Study

Between-subjects contrasts	Similarity ratio	
	SR = 1	SR = 9
Order 1		
Labels-plus-training vs. no-labels	0.80	0.44
Labels-minus-training vs. no-labels	0.60	0.83
Order 2		
Labels-plus-training vs. no-labels	1.24	0.92
Labels-minus-training vs. no-labels	0.53	0.55

Table 1), how can the following conclusions be drawn? A. “labels do not generally affect children’s perceptual similarity judgments” (p. 890); B. “labels influence perceptual similarity primarily under the adverse order of conditions” (p. 894); or C. “children’s reliance on labels to make similarity judgments appears to be attributable to flaws in the methodological approaches used in prior studies” (p. 890). Our answer is these conclusions are inconsistent with Noles and Gelman’s (2012) own data and are therefore misleading.

Finally, despite their claims to the contrary, Noles and Gelman’s (2012) data perfectly replicate Sloutsky and Fisher’s (2004) findings even in the most critical Order 1. To demonstrate that, we used Noles and Gelman’s data to estimate the values of  $w$ —the attentional weight of the label. This was easy to do, given Noles and Gelman’s similarity judgment data in both label and no-label conditions (we just divided the similarity ratio observed in the label condition by the similarity ratio in the no-label condition). Estimates of  $w$  for the label-plus-training condition in Order 1 (this is the most critical condition, according to Noles and Gelman’s argument) are presented in Table 2. Also in the table are estimates of  $w$  used in Sloutsky and Fisher (these were derived from earlier work by Sloutsky & Lo, 1999). Having derived  $w$  values from Noles and Gelman’s data, we could examine how well these parameter values predict findings reported by Sloutsky and Fisher (Experiment 1). To obtain these predictions, we plugged estimates of  $w$  derived from Noles and Gelman data into SINC (Similarity, Induction, Categorization, and Naming) model, specifically Equation 5 in Sloutsky and Fisher (2004):

$$P(B) = \frac{1}{1 + \frac{WS^y}{S^x}}$$

where  $S^y/S^x$  is the similarity ratio in the no-label condition. Given that values of all similarity ratios used by Sloutsky and Fisher are known from extensive calibration data collected by Sloutsky and Fisher, we could obtain predicted probabilities of choosing the test item sharing the label with the target as looking more similar to the target. We could then compare these predicted values with the values reported by Sloutsky and Fisher. These predicted and observed data are presented in Figure 1.

As is shown in Table 2, parameter values derived from Noles and Gelman’s (2012) data are very close to the value derived from Sloutsky and Lo (1999) data and used in Sloutsky and Fisher (2004). Furthermore, as can be seen in Figure 1, both parameter values derived from Noles and Gelman accurately predict Sloutsky and Fisher’s data on similarity judgment in the label condition, which means nothing short of a convincing replication of Sloutsky and Fisher’s findings by Noles and Gelman. Also note that all

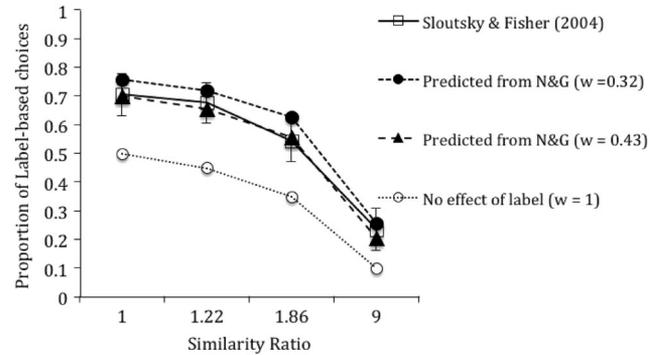


Figure 1. Similarity judgment (a) reported by Sloutsky and Fisher (2004), (b) predicted by using the values of the label parameter derived from Noles and Gelman’s (2012; N&G) data (see Table 2), and (c) when labels have no effect on similarity judgment.

these situations differ markedly from what should have been observed under  $w = 1$ —a situation when labels have no effect on similarity judgment. Of course, if a design uses only two levels of similarity ratio (such as Noles and Gelman’s design), it is less likely to capture these differences than designs that include multiple levels (such as Sloutsky and Fisher’s design).

Therefore, it seems that three firm conclusions follow from Noles and Gelman’s (2012) experiments: (a) Similarity ratio affects similarity judgment of young children, which indicates that Sloutsky and Fisher (2004) adequately calibrated their stimuli; (b) the order of presentation affects similarity judgment of young children (and the order may moderate the effect of similarity ratio), which is a new finding (recall that Sloutsky and Fisher randomized the order of presentation); and (c) labeling affects similarity judgment of young children, even when precautions are taken to ensure proper understanding of the similarity question (which is a perfect replication of Sloutsky and Fisher). The former two effects have little to do with the debate about the role of labels in similarity judgment because these effects transpire for both label and no-label conditions (with no interaction with labeling). The third effect relates directly to the debate: By replicating Sloutsky and Fisher it presents additional evidence that labels affect similarity judgment of young children. Such a replication, although not being a novel finding, is an important contribution to the debate about the effects of labels on similarity because it comes from the lab disputing such effects.

## References

- Deng, W., & Sloutsky, V. M. (2012). Carrot-eaters and moving heads: Salient features provide greater support for inductive inference than category labels. *Psychological Science*, *23*, 178–186. doi:10.1177/0956797611429133
- Fisher, A. V. (2010). What’s in the name? Or how rocks and stones are different from dogs and puppies. *Journal of Experimental Child Psychology*, *105*, 198–212. doi:10.1016/j.jecp.2009.11.001
- Fisher, A. V., Matlen, B., & Godwin, K. E. (2011). Semantic similarity of labels and inductive generalization: Taking a second look. *Cognition*, *118*, 432–438. doi:10.1016/j.cognition.2010.12.008
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York, NY: Oxford University Press.

Table 2

Attentional Weights of Label ( $w$ ) Derived From Noles and Gelman (2012) Data and Used in Sloutsky and Fisher (2004)

Condition	Value of $w$
Noles and Gelman (2012): Order 1/Label-plus-training/SR = 9	0.32
Noles and Gelman (2012): Order 1/Label-plus-training/SR = 1	0.43
Sloutsky and Fisher (2004; for all levels of similarity ratio)	0.35

- Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology, 26*, 796–804. doi:10.1037/0012-1649.26.5.796
- Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences* (3rd. ed.). Pacific Grove, CA: Brooks/Cole.
- Noles, N. S., & Gelman, S. A. (2012). Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches. *Developmental Psychology, 48*, 890–896. doi:10.1037/a0026075
- Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*, 246–251. doi:10.1016/S1364-6613(03)00109-8
- Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science: A Multidisciplinary Journal, 34*, 1244–1286. doi:10.1111/j.1551-6709.2010.01129.x
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133*, 166–188. doi:10.1037/0096-3445.133.2.166
- Sloutsky, V. M., & Fisher, A. V. (2005). Similarity, Induction, Naming, and Categorization (SINC): Generalization or verbal inductive reasoning? Response to Heit and Hayes. *Journal of Experimental Psychology: General, 134*, 606–611. doi:10.1037/0096-3445.134.4.606
- Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology, 111*, 65–86. doi:10.1016/j.jecp.2011.07.007
- Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1. Linguistic labels and the development of similarity judgment. *Developmental Psychology, 35*, 1478–1492. doi:10.1037/0012-1649.35.6.1478
- Sloutsky, V. M., Lo, Y.-F., & Fisher, A. (2001). How much does a shared name make things similar? Linguistic labels and the development of inductive inference. *Child Development, 72*, 1695–1709. doi:10.1111/1467-8624.00373
- Waxman, S. R., & Markow, D. B. (1995). Words as invitation to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302. <http://dx.doi.org/10.1016/j.bbr.2011.03.031>. doi:10.1006/cogp.1995.1016

Received January 14, 2012

Accepted January 19, 2012 ■