

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Language Shifts the Representation of Sounds in Time: From Auditory Individuals to Auditory Ensembles

Permalink

<https://escholarship.org/uc/item/9329q9p8>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

Ongchoco, Julia
Knowlton, Tyler
Papafragou, Anna

Publication Date

2023

Peer reviewed

Language Shifts the Representation of Sounds in Time: From Auditory Individuals to Auditory Ensembles

Julia Ongchoco (juliaong@sas.upenn.edu)

MindCORE, 3740 Hamilton Walk
Philadelphia, PA 19104 USA

Tyler Knowlton (tzknowlt@upenn.edu)

MindCORE, 3740 Hamilton Walk
Philadelphia, PA 19104 USA

Anna Papafragou (anna4@sas.upenn.edu)

Department of Linguistics, 3401 Walnut Street
Philadelphia, PA 19104 USA

Abstract

Objects can either be represented as independent individuals (“object-files”) or as members of a collection (an “ensemble”). Work over the past 40 years has explored these representational systems, largely in the visual domain. Far less is known about auditory objects. Here, we show that a property characteristic of visual object representation – that it can be modulated by linguistic framing – also applies to auditory objects. In particular, we show that using the expression “each sound” versus “every sound” can bias auditory object construal in the same way that using “each circle” versus “every circle” can bias visual object construal. These findings support the idea that object-files and ensembles are not limited to the visual domain, but are representational formats found more generally throughout cognition.

Keywords: Auditory Objects, Ensemble Representation, Language, Psycholinguistics, Quantification, Linguistic Framing

Introduction

Objects in the world can be mentally represented in at least two ways: as a series of independent individuals or as members of a single group. On the individual side, humans (as well as other animals) are able to encode objects as “object-files” (e.g., Kahneman & Treisman, 1984; Kahneman, Treisman, & Gibbs, 1992). These representations consist of an index that serves as an anchor for associated properties (e.g., size, color, orientation). Over the last four decades, much has been learned about this system of individual object representation, its properties, and how it develops from infancy to adulthood (for reviews, see Carey, 2009 and Green & Quilty-Dunn, 2021).

More recently, a parallel literature has emerged exploring group representations. Of particular relevance here, Arieli (2001) introduced the notion of an “ensemble” (for reviews, see Alvarez, 2011; Haberman & Whitney, 2012; Whitney & Yamanashi Leib, 2018). Instead of representing individuals as such, ensemble representation (sometimes referred to as ensemble processing, perception, or coding) allows for representing multiple objects simultaneously by abstracting away from the individuals and encoding the collection in

terms of summary properties (e.g., average size, average hue, average orientation). These summary properties include both measures of central tendency (e.g., average hue) and measures of variance (e.g., color diversity). Perhaps most interestingly, a common theme in the ensemble literature has been the counterintuitive idea that such ensemble summary properties can be computed even in the absence of precise information about the individual items that constitute the ensemble (e.g., Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008; Haberman & Whitney, 2011; Im & Halberda, 2013; Ward, Bear, & Scholl, 2016).

But despite the plethora of research on these two representational systems – object-files and ensembles – the bulk of the findings pertain to the visual modality (*cf.* work on objects in the haptic domain: Riggs, Ferrand, Lancelin, Fryziel, Dumur, & Simpson, 2006; Plaisier, Tiest, & Kappers, 2009). Comparatively little is known about the extent to which auditory objects can be represented with the same two systems under similar conditions.

To be sure, some principles that apply to visual object individuation are likely to have parallels in the auditory domain. For example, attention can be captured by hearing one’s name mentioned even in the presence of background chatter, just as attention can be captured by the appearance of a new visual object (e.g., Bregman, 1994; Shinn-Cunningham, 2008). And there are likewise parallels between visual and auditory ensembles. Piazza, Sweeny, Wessel, Silver, and Whitney (2013) offer evidence that participants can represent a series of tones in terms of their summary properties, namely their average frequency. And, as is characteristic of ensemble coding, participants could do so even when their knowledge of individual tones in the group was degraded.

This finding is potentially surprising, as auditory objects are temporally discontinuous, meaning the extraction algorithm(s) deployed to get from the distal stimulus to the ensemble representation must be quite different from the analogous algorithm(s) deployed in vision. Put another way, however ensemble statistics are calculated for sequentially

encountered objects is likely different in kind from how those same statistics are calculated for simultaneously displayed objects (of the kind often used in visual experiments). Nonetheless, these results suggest that there are amodal ensemble representations, which represent collections by way of their summary properties.

Building on this initial work, we investigate how flexible these sorts of representations are by asking whether the construal of auditory objects – as individual object-files or as an ensemble collection – can be modulated by linguistic framing. In particular, we hypothesize that describing a series of tones as “each sound” will encourage individual construals of the tones whereas describing the same tones as “every sound” will encourage ensemble coding. This hypothesis aligns with recent work on linguistic framing in the visual domain, which finds that describing some circles with expressions like “each circle is green” encourages adults and children to represent them as independent individuals, but describing them instead with expressions like “every circle is green” (or “all the circles are green”) instead encourages participants to treat them as an ensemble collection.

Linguistic Framing & Visual Objects

As noted above, our main question is whether linguistic framing can affect auditory objects in the same way that it can affect visual objects. Of course, linguistic framing effects are more widespread throughout cognition. Perhaps the most well-known cases come from experiments on decision-making, which compare, for example, describing a medical procedure in which “30 of the last 50 operations have been successful” to one in which “20 of the last 50 operations have failed”. Despite the mathematical success rates being identical, the language used seems to matter: participants are often shown to prefer the procedure described in terms of its success rate over the one described in terms of failure (e.g., Duchon, Dunegan, & Barton, 1989; Kahneman & Tversky, 1979; Geurts, 2013).

More generally, describing the same state of the world with different expressions can alter one’s perspective. One might have different thoughts about a particular event if that event is described as a “chasing” instead of a “fleeing”, for example (Gleitman, 1990). Similarly, choice of expression has been shown to impact moral judgments. In particular, difference in use of agentive or non-agentive language (“She ignited the carpet” vs. “The carpet ignited”) can influence the way people assign blame and fault: the more agentive the expression, the more blame assumed (e.g., Fausey & Boroditsky, 2010).

Moving away from reasoning and morality, comparatively lower-level linguistic framing effects have been found in the visual domain. To take a recent example, Knowlton, Hunter, Odic, Wellwood, Halberda, Pietroski, and Lidz (2021) report that when participants are asked to create images of blue and yellow dots in which “most of the dots are blue”, both adults and children prefer images of blue and non-blue dots clustered together. But when instead given the

logically-equivalent instruction “more of the dots are blue”, participants prefer spatially separated blue and non-blue dots. Moreover, when shown the very same image of blue and yellow dots, only those who heard “more of the dots are blue” attended to the yellows; those who instead heard “most of the dots are blue” attended only to the blues and the superset of all dots. As in the reasoning example above, the choice of expression used in posing the question matters.

In this paper, we aim to extend linguistic framing effects to the construal of auditory objects in service of exploring the nature of auditory ensembles. We leverage recent work on a linguistic framing effect in the visual domain: the English quantifiers “each” and “every”. Though these two expressions are obviously similar (e.g., “every note was sung perfectly” is true just in case “each note was sung perfectly” is true), their meanings nonetheless differ. In particular, “each” has long been known to have a ‘more individualistic’ meaning than its counterpart “every” (e.g., Vendler, 1962; Beghelli & Stowell, 1997; Tunstall, 1998). To see this, consider questions like those in (1).

- (1) a. Which song did each student sing?
- b. Which song did every student sing?

The “each”-variant in (1a) raises the possibility that each student sang a different song. As such, it might be answered with a list: “Tina sang *Take Me Out To the Ball Game*, Harry sang *High Hopes*, Lisa sang *Let It Be...*”. But the “every”-variant in (1b) seems to be asking about which one song all the students sang, and, accordingly, (1b) resists the sort of list-based response that (1a) invites.

Partly in response to data like these, Knowlton (2021) and Knowlton, Pietroski, Halberda, and Lidz (2022) propose that “each” and “every” have formally distinct concepts of universal quantification as their meanings. The gist of the proposal is that the meaning of “every” calls for grouping the things quantified over (e.g., the students) whereas the meaning of “each” calls for treating the things quantified over as independent individuals. Put another way, the claim is that the meaning of “every student” has a constituent that corresponds to the plurality “the students”, whereas the meaning of “each student” lacks any such constituent.

A large portion of the evidence for this view comes from linguistic framing effects in the visual domain. In these sorts of experiments, the main result is that participants group together visually presented objects as an ensemble collection if those objects are quantified over with “every”, but instead deploy their object individuation system when the very same objects are quantified over with “each”. In one case, for example, participants were shown images of different-colored circles and either asked to verify sentences like “each circle is green” or “every circle is green”. Participants in both conditions were able to complete the task without issue. But after giving their answer, they were asked a memory question that probed a group-property, like “how many circles were there?”. Participants in the “every” condition performed better than those in the “each”

condition (Knowlton, Pietroski, Halberda, and Lidz, 2022). In a separate experiment, participants were instead asked a question that probed an individual-property, like “did this particular circle change color?”. Here, those in the “each” condition performed better than those in the “every” condition (Knowlton, 2021).

Evidence from pragmatic use has also been marshaled in support of the above view. Knowlton, Trueswell, and Papafragou (2022) show that as the number of things quantified over increases, participants prefer to use “every” instead of “each”. This is explained by appealing to a well-known difference between object-files and ensembles: only object-files are subject to a strict working memory limit (e.g., Feigenson & Carey, 2005). Furthermore, Knowlton and Gomes (2022) demonstrate that parents are sensitive to this distinction in their speech to children: they are more likely to use “each” when talking about small numbers of things and “every” (or “all”) when quantifying over more than three objects.

In sum, when visually presented objects are described with the expression “each object”, participants often deploy their object-file system to represent those objects. But when the same visual objects are described with the expression “every object”, participants instead are biased to represent them with their ensemble system. These differences in entity construal – as independent individuals versus as an ensemble collection – influence what sorts of properties participants encode and recall, as well as the conditions under which participants prefer to use either expression.

In what follows, we ask whether these sorts of ‘low-level linguistic framing effects’ likewise extend to the auditory domain. Namely, we explore whether describing a series of tones with “each” versus “every” changes the representational system that participants are biased to deploy to represent those tones. We do so by probing the information they encode and recall after evaluating sentences with “each” or “every”. Participants who first evaluate a sentence with “each sound” are predicted to be better at a subsequent change detection task (probing an individual property) than participants who instead encounter and evaluate expressions with “every sound”.

We report two experiments that provide initial evidence of this effect. The first was conducted on undergraduate participants and used a subjective measure of attention; the second was conducted on a more heterogeneous online sample and included the addition of attentional catch trials. In both cases, we find a slight advantage for “each” over “every” when participants are asked whether the average tone was among three initial tones they heard.

Experiment 1

Experiment 1 tests the prediction that describing sounds using the quantifiers “each” or “every” will impact whether those sounds are represented as independent object-files or as members of an ensemble collection. Participants were first asked to evaluate an introductory sentence that used either “each” or “every” (e.g., “each sound is pleasant”)

with respect to three sounds. Then, they were asked to determine whether a fourth ‘test’ sound was one of the initial three tones that they just heard or not.

Our main prediction, in line with the visual work discussed above, is that evaluating an initial sentence with “each” should bias participants to treat the sounds as independent objects, in turn leading to superior performance on the test questions when compared to “every”, which should instead bias participants to group the sounds as an ensemble, in turn leading to worse performance. This is because the test questions used here are questions that probe an individual property – the frequency of a particular sound – and not a group property (like the average frequency or the number of tones played).

We were especially interested in trials in which the test tone was not present among the initial three tones, but was the average frequency of the initial three. Given that ensembles represent collections in terms of averages, we expected higher rates of false alarms on these average trials in the “every” condition.

Participants Two-hundred participants were recruited from the University of Pennsylvania Department of Psychology subject pool. This large number of participants was intended to roughly mirror the number of trials from Piazza et al. (2013) – the closest comparable study – which used fewer subjects but a larger number of trials. Each participant gave informed consent prior to participating (in accordance with the University of Pennsylvania IRB) and was compensated with course credit.

Apparatus Participants completed the experiment in a web browser, with stimulus presentation and data collection controlled via custom software written using a combination of HTML, CSS, JavaScript, PHP, and jsPsych libraries (De Leeuw, 2015). To discourage multitasking, participants were required to complete the experiment with their browser in fullscreen mode.

Stimuli Each sound lasted 250ms. The sounds used ranged from 65.41Hz to 1975.53Hz, in semitone steps (in musical notation, C2 to B6), and were rendered using the built-in Javascript audio function. For each trial, a frequency was sampled from this range to serve as the ‘middle’ tone. Following previous work (Piazza et al., 2013), a tone triplet was then created by subtracting 7 semitones from this middle tone (to create the second tone in the triplet) and then adding 7 semitones from the middle tone (to create the final tone in the triplet). This triplet constituted the initial three tones that served as the basis for sentence evaluation (i.e., these were the tones on which participants based their initial judgments for subjective sentences like “each sound is beautiful”). For each set of three initial tones, two types of ‘test’ tones were created. ‘Wrong’ tones were selected to be 12 semitones from the middle tone from the initial set, and ‘Average’ tones were determined by obtaining the average frequency of all the tones in the triplet (with an average

difference of 31 Hz from the middle tone across trials). In the present experiment, the linear average was used (though in future work, we plan to compare this to the logarithmic average).

Procedure & Design At the start of the experiment, participants were asked to adjust their volume to a comfortable level. In each trial, participants first read a sentence describing a subjective opinion about three sounds (e.g. “Each sound is pleasant”; “Every sound is calm”). These sentences used both positively and negatively valenced adjectives (“pleasant”, “calm”, “serene”, “positive”, “uplifting”, “enjoyable”, “dull”, “creepy”, “odd”, “annoying”, “shrill”, and “unnerving”).

These subjective opinion sentences either indicated universal quantification with “Each” or with “Every” (our two main conditions of interest). Participants were randomly assigned to one of the two conditions at the start of the experiment. We avoided a within-subjects design as past work has found order effects with these sorts of tasks (e.g., it is often hard to encourage participants to shift away from a ‘superior’ strategy to an ‘inferior’ one).

After seven seconds, the sentence disappeared, and the tones from the triplet played sequentially (for 250ms each). Immediately following the presentation of sounds, the subjective opinion sentence appeared again, and participants were asked to report whether they agreed or disagreed with the opinion by pressing either the “A” key for “Yes” or the “L” key for “No.” There was a delay of 500ms after their keypress before a follow-up memory question appeared: “Memory Test: Did you just hear this sound? Get Ready.” After three seconds, the follow-up memory question disappeared, and the test sound played. This test sound could be one of three types: (1) a sound that was one of the three sounds that were initially played (‘Present’), (2) a sound that was the average of the three sounds but not a part of the triplet (‘Average’), or (3) a sound that was above or below the boundaries of initial triplet (‘Foil’; i.e., a sound selected to be 12 semitones away from the middle tone from the initial triplet). Present trials could consist of the first, second, or last tone in the triplet (this was counterbalanced through multiple stimuli lists).

Participants were given seven seconds to report whether the sound had played earlier, by again pressing the “A” key for “Yes” or the “L” key for “No.” Participants completed 36 trials (18 Present trials, 9 Average trials, and 9 Foil trials), presented in a random order per participant.

Finally, at the end of the experiment, participants completed a debriefing questionnaire, which included questions asking about their musical training experience, and how well they paid attention during the task (on a scale of 1-100, with 1 being “Very distracted, not at all”, and 100 being “Very focused throughout the experiment”).

Results Overall accuracy on follow-up memory questions for participants who reported an attention level below the median (80%) was significantly lower than it was for

participants who reported an attention level above the median, 59% vs. 68% (Welch's t-test: $t(182.12)=4.86$, $p<.001$). Given the general difficulty of the task, we opted to analyze and report only the results from the highest-engagement participants, who reported their attention to be 80% and above. See Experiment 2 for evidence that this cutoff captures attentiveness during the task. This left 116 participants.

Performance on the three test sound types – Average, Foil, and Present – is shown in Figure 1. There was no significant difference between overall performance in the two conditions (Each: 69%; Every: 67%; $t(111.4)=0.71$, $p=.45$), but as Figure 1 shows, there was a difference based on the type of sound played at test. In particular, on the foil trials, participants who initially evaluated “each”-sentences performed better than those who listened to the same tones but initially evaluated “every”-sentences.

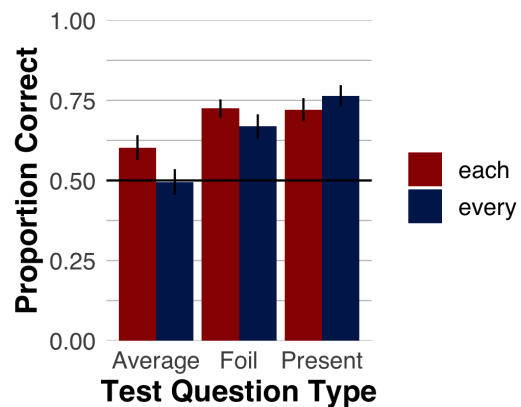


Figure 1: Average performance separated by question type: ‘Average’ questions probed the average of the initial triplet (which was not present in that initial set), ‘Foil’ questions probed a tone outside of the range of the initial triplet, ‘Present’ questions probed a tone that was present in the initial triplet. Error bars represent standard error.

To verify this statistically, we fit a mixed effects model with fixed effects for quantifier (“Each” versus “Every”), question type (Average, Foil, or Present), and their interaction as well as random intercepts for participant. This model is a better fit than a comparable model without the interaction term (only main effects; $\chi^2(2)=22.03$, $p<.001$) and it yields a significant interaction between quantifier type and question type ($\beta=.16$ [95% CI .10 to .21], $z=2.99$, $p<.01$). This interaction seems to be largely driven by the difference in performance in the Average condition. Post-hoc t-tests show a significant difference between the Each and Every condition for questions that probe the average ($t(1020.1)=3.56$, $p<.001$), but only a marginally significant difference between Each and Every for questions that probe a sound outside of the initially-presented range (Foil condition: $t(1014.5)=1.95$, $p=.052$). Moreover, only participants who first evaluated sentences with “Each” performed significantly above chance in the Average

condition (Each: $t(482)=4.6$, $p<.001$; Every: $t(551)=-.34$, $p=.63$). So while questions probing the average were harder overall than questions probing a foil outside the range of the initial trio or a sound that was present (significant main effect of question type: $\beta=-.54$ [95% CI $-.59$ to $-.48$], $z=10.23$, $p<.001$), participants who initially evaluated “each”-sentences performed better on follow-up questions probing the average tone than those who listened to the same tones but initially evaluated “every”-sentences. This pattern of performance was observed for 11 out of the 12 adjectives used.

These results suggest that participants may have been spontaneously extracting the average tone in the Every condition, thereby making them more likely to incorrectly answer “yes” when they were given a follow-up memory question asking whether that average was a previously-heard tone. This finding is well-explained if “every” biases participants to rely on ensemble representations of sounds, as has been previously suggested for visually presented objects. Since ensemble representations rely on estimations of central tendency, the average frequency constitutes a particularly attractive lure, despite the fact that it was not one of the sounds initially presented to participants. But given that ensemble representations likely do include measures of the range, it makes sense that our Foil trials (which contained a sound outside of the initial range) were not as difficult for those in the Every condition.

Still, this effect is surprising, in part because the only difference between conditions was the quantifier used. And sentences with “each” and “every” are logically equivalent. That is, each sound is creepy/enjoyable/unnerving/etc. if and only if every sound is creepy/enjoyable/unnerving/etc. Indeed, participants’ rate of answering “yes” to the initial subjective questions did not differ by condition (38.76% for Each vs. 41.42% for Every; $t(113)=0.68$, $p=0.49$).

Experiment 2

The results from experiment 1 suggest that quantifying over the tones with “each” encourages subjects to create object-file representations of the individual sounds. These representations, in turn, make participants better able to pinpoint subtle differences in the tones, compared to the ensemble representations of tones encouraged by “every”. However, the overall difficulty of the task and the importance of participants paying attention to the wording of the initial questions led us to exclude participants based on a subjective report of their attention level. In Experiment 2, a more objective measure of attention was used instead. In particular, we added ‘catch’ trials after every eight experimental trials. These catch trials asked participants to recall the adjective in the sentence they were meant to evaluate. We expect to find the same result if inattentive participants are excluded on this basis as opposed to on the basis of their subjective attention reports.

Methods Experiment 2 was identical to Experiment 1, with a few exceptions. First, Prolific (prolific.co) was used to recruit participants instead of the undergraduate subject pool. Each participant gave informed consent (in accordance with the University of Pennsylvania IRB) and was compensated with \$2.50 (\$10 per hour). Given the increased heterogeneity of this online sample, the numbers of recruited subjects and trials were both increased. The number of participants was doubled, to 400 in total, and 12 more trials were added (6 Test trials, 3 Average trials, and 3 Foil trials), for a total of 48 trials per participant.

Second, given that the Foil trials were easier than the Average trials in Experiment 1, we slightly increased the difficulty for some of the Foil trials in Experiment 2. Half of the group received the same tones that were used in Experiment 1 (where Foil trials were generated by selecting sounds that were 12 semitones away from the middle tone). But the other half of the group received Foil tones that added or subtracted 9 semitones from the middle tone, rather than 12.

Finally, and most importantly, catch trials were added that asked subjects to “select the adjective that was used to describe the tones.” This statement was followed by four options of synonyms to the adjective in the last trial (evenly distributed across Present, Average, and Foil test trials).

Results We initially excluded any participants who did not complete the entire experiment ($n=14$) and any who took longer than 30 minutes to complete the experiment ($n=28$), as this was nearly double the median time of the full sample (Every condition: 15.8 minutes; Each condition: 15.7 minutes). We also excluded participants who performed at 50% or below – fewer than 4/6 correct answers – on the attention check trials ($n=42$). This left 316 participants. Removal based on explicit attention checks resulted in nearly every single participant having a subjective attention score at or above 80% (only a single participant left in the sample reported lower than 80%), supporting its use as an exclusion criterion in Experiment 1.

Participants’ performance on the three test sound types is shown in Figure 2. Similar to Experiment 1, there was not a significant difference between overall performance in the two conditions (Each: 69%; Every: 67%; $t(313.9)=1.89$, $p=.06$). But as Figure 2 shows, there was a significant interaction between quantifier type (“Each” versus “Every”) and question type (Average, Foil, or Present) such that performance was better in the Each condition (comparison against no-interaction model: $\chi^2(2)=17.95$, $p<.001$; significant interaction: $\beta=.07$ [95% CI $.04$ to $.10$], $z=2.42$, $p<.05$). So while questions probing the average were far harder than questions probing foil tones or present tones (significant main effect of question type: $\beta=-1.33$ [95% CI -1.36 to -1.30], $z=44.00$, $p<.001$), participants who initially evaluated “each”-sentences produced fewer ‘false alarms’ than those who listened to the same tones but initially evaluated “every”-sentences.

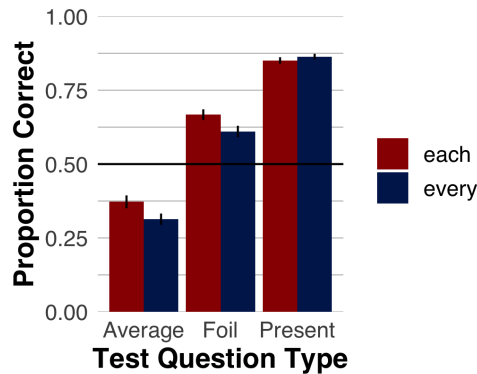


Figure 2: Average performance separated by question type: ‘Average’ questions probed the average of the initial triplet (which was not present in that initial set), ‘Foil’ questions probed a tone outside of the range of the initial triplet, ‘Present’ questions probed a tone that was present in the initial triplet. Error bars represent standard error.

Notably, performance in both the Each and Every conditions in Experiment 2 was numerically lower than in Experiment 1. We suspect this may be due to differences between the subject pools or to the increased number of trials. We hope to explore why this might have been the case in future variants of the task. Nonetheless, it is reassuring to see that some effect of quantifier is present despite participants’ performance being closer to floor.

General Discussion

Here, we explored how the choice between two seemingly similar quantificational words can invite the use of two distinct representational systems in the auditory domain. We asked whether indicating universal quantification with “each” versus “every” can play a role in determining whether the auditory objects quantified over are represented as object-files or as members of an ensemble. And in two experiments, we found initial evidence that quantifying over a sequence of three sounds with “every” instead of “each” altered how likely participants were to incorrectly report that a tone was part of that sequence.

This effect was largely (though not entirely) limited to the cases in which average tones were presented as lures, which can be explained by a feature of the ensemble representations that “every” invites. In particular, ensembles encode collections of objects in terms of summary properties, like average hue or average frequency. So participants were more likely to mistake averages as tones they had heard if they were pushed (by the use of that quantifier) to treat those tones as members of an ensemble. But given that ensemble representations do offer measures of range in addition to measures of central tendency, it seems reasonable that the effect was not as pronounced on our ‘Foil’ trials, which presented tones above or below the range of the initial trio. Still, the representation of the average seemed to have interfered with those cases as well.

Importantly though, choice of quantifier did not seem to alter performance on the initial subjective opinion questions (e.g., “is each/every sound annoying?”). Thus, as argued in the visual work cited above, the difference in meaning between the quantifiers “each” and “every” is subtle enough to change the way participants represent the things quantified over, without suggesting distinct informational content. Put another way, the complex concepts constructed in response to hearing “each sound is creepy” and “every sound is creepy” are judged as true in the same situations, but despite this logical equivalence, these meanings have distinct formats, which bias participants to deploy distinct representational systems. Namely, the thought constructed in response to hearing “every sound” contains a constituent corresponding to the plurality “the sounds”. This group representation encourages participants to extract ensemble summary statistics (e.g., average tone) during sentence evaluation, which in turn makes those summary statistics available for use in downstream decisions. In this case, the availability of the ensemble representation led to a worse performance (an increase in false alarms). In contrast, when instead primed with “each sound”, the thought constructed has no part corresponding to the plurality “the sounds”, so participants were more likely to avoid extracting ensemble summary statistics like the average frequency.

Of course, this suggests that in other tasks “every sound” should lead to superior performance. For instance, if participants were asked a follow-up question about whether a fourth tone was above or below *the average of the previous three*, we predict they will be more likely to succeed if primed with “every” than if primed with “each”. In ongoing work, we are conducting this variant of the task.

For now, it is worth noting that the effect observed is small. We suspect this has to do with the extreme difficulty of the trials probing the average, as evidenced by the below-chance performance in both conditions in Experiment 2. The average tones used here tended to be quite close to tones that did occur in the initial trio. In future versions of the task, we plan to decrease the difficulty of these trials, in the hopes of magnifying the effect. Nonetheless, using “each” versus “every” does seem to give rise to a difference.

This finding builds on previous work in three ways. In terms of the impact of language on object construal, our findings extend the linguistic framing effects of “each” and “every” from the visual to the auditory domain; from static visual displays to sounds that unfold over time. This supports the generality of those proposals (i.e., they are not limited to language-*vision* interfaces). In terms of the object-file versus ensemble distinction more broadly, our results lend support to the idea that ensemble representations are operative in audition. Finally, these results speak to the malleability of object construal. Though object-files and ensembles can obviously be triggered by bottom-up properties of stimuli displays – number of objects, homogeneity of objects, spatial layout – the deployment of these two representational systems can also be modulated by linguistic framing.

Acknowledgments

We thank Joan Ongchoco for assistance with experiment coding and analysis, Rebecca Peng for help with participant testing, and members of the Penn Language and Cognition Lab for helpful discussion. We also thank MindCORE for generously funding this project as part of the 'Representations of Quantity in Language and Cognition' collaborative research grant.

References

- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences*, 15(3), 122-131.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12(2), 157-162.
- Beghelli, F. & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In Szabolcsi, A. (ed.) *Ways of Scope Taking*, 71-107. Springer.
- Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.
- Carey, S. (2009). *The origin of concepts*. Oxford Series in Cognitive Development. Oxford University Press.
- Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia*, 46(11), 2861-2864.
- De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1-12.
- Duchon, D., Dunegan, K. J., & Barton, S. L. (1989). Framing the problem and making decisions: the facts are not enough. *IEEE Transactions on Engineering Management*, 36(1), 25-27.
- Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, 17(5), 644-650.
- Feigenson, L. & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition*, 97(3), 295-313.
- Geurts, B. (2013). Alternatives in framing and decision making. *Mind & Language*, 28(1), 1-19.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3-55.
- Green, E. J. & Quilty-Dunn, J. (2021). What is an object file? *The British Journal for the Philosophy of Science*, 72(3), 665-699. *The British Journal for the Philosophy of Science* 2021 72:3, 665-699.
- Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. From perception to consciousness: Searching with Anne Treisman, 339-349.
- Haberman, J., & Whitney, D. (2011). Efficient summary statistical representation when change localization fails. *Psychonomic Bulletin & Review*, 18, 855-859.
- Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, 75, 278-286.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: an analysis of decisions under risk. *Econometrica*, 47, 263-291.
- Kahneman, D. & Treisman, A. (1984). Changing views of attention and automaticity in varieties of attention. In Parasuraman, R. & Davies, D. R. (eds.), *Varieties of Attention*, 29-61. Academic Press.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175-219.
- Knowlton, T. (2021). *The psycho-logic of universal quantifiers*. University of Maryland dissertation.
- Knowlton, T., Hunter, T., Odic, D., Wellwood, A., Halberda, J., Pietroski, P., & Lidz, J. (2021). Linguistic meanings as cognitive instructions. *Annals of the New York Academy of Sciences*, 1, 134-144.
- Knowlton, T., Pietroski, P., Halberda, J., & Lidz, J. (2022). The mental representation of universal quantifiers. *Linguistics & Philosophy*, 45, 911-941.
- Knowlton, T., Trueswell, J., & Papafragou, A. (2022). A Mentalistic Semantics Explains "Each" and "Every" Quantifier Use. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44).
- Knowlton, T., & Gomes, V. (2022). Linguistic and non-linguistic cues to acquiring the strong distributivity of each. *Proceedings of the Linguistic Society of America*, 7(1), 5236.
- Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, 24(8), 1389-1397.
- Plaisier, M. A., Tiest, W. M. B., & Kappers, A. M. (2009). One, two, three, many—subitizing in active touch. *Acta Psychologica*, 131(2), 163-170.
- Riggs, K. J., Ferrand, L., Lancelin, D., Fryziel, L., Dumur, G., & Simpson, A. (2006). Subitizing in tactile perception. *Psychological Science*, 17(4), 271-272.
- Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12, 182-186.
- Tunstall, S. (1998). *The interpretation of quantifiers: Semantics and processing*. University of Massachusetts Amherst dissertation.
- Vendler, Z. (1962). Each and every, any and all. *Mind*, 71(282), 145-160.
- Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition*, 152, 78-86.
- Whitney, D. & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105-129.