



Enumeration in time is irresistibly event-based

Joan Danielle K. Ongchoco¹ · Brian J. Scholl¹

Published online: 21 January 2020
© The Psychonomic Society, Inc. 2020

Abstract

One of the most fundamental questions that can be asked about any process concerns the underlying units over which it operates. And this is true not just for artificial processes (such as functions in a computer program that only take specific kinds of arguments) but for mental processes. Over what units does the process of enumeration operate? Recent work has demonstrated that in visuospatial arrays, these units are often irresistibly discrete objects. When enumerating the number of discs in a display, for example, observers underestimate to a greater degree when the discs are spatially segmented (e.g., by connecting pairs of discs with lines): you try to enumerate discs, but your mind can't help enumerating dumbbells. This phenomenon has previously been limited to static displays, but of course our experience of the world is inherently dynamic. Is enumeration in time similarly based on discrete events? To find out, we had observers enumerate the number of notes in quick musical sequences. Observers underestimated to a greater degree when the notes were temporally segmented (into discrete musical phrases, based on pitch-range shifts), even while carefully controlling for both duration and the overall range and heterogeneity of pitches. Observers tried to enumerate notes, but their minds couldn't help enumerating musical phrases – since those are the events they experienced. These results thus demonstrate how discrete events are prominent in our mental lives, and how the units that constitute discrete events are not entirely under our conscious, intentional control.

Keywords Event segmentation · Number estimation · Auditory streaming · Auditory perception

Introduction

If you want to understand how a function in a computer program works, one of the first and most important things you'll need to know is just what kinds of arguments or messages that function takes as input. (Images? Text strings? Floating-point numbers? Etc.) So too for mental processes – and accordingly a great deal of work (and controversy) has involved specifying the underlying units of various aspects of cognition.

In research on visual attention, for example, we must ask: To what can we attend in the first place? Features? Spatial regions? In fact, previous work has demonstrated that we often attend irresistibly to discrete objects – such that we can better discriminate two features when they lie on the same

object versus two distinct objects, even while controlling for spatial distance (for a review, see Scholl, 2001). And critically, we don't have full control over what counts as an object; rather, we must attend to what our visual system automatically treats as objects (so that, for example, if we are asked to attentionally track particular ends of dumbbells, we often fail as attention automatically selects the entire dumbbells; Scholl et al., 2001).

What counts (in space)?

In the present study, we asked about the underlying units of mental enumeration. Our “number sense,” of course, is a key part of our “core knowledge” about the world (e.g., Spelke & Kinzler, 2007), and numerical representations interact in rich ways with many other aspects of our minds, from language to spatial representation (for a review, see Dehaene, 1997). In addition, the ability to extract numerosity from perceptual input occurs both across the animal kingdom (for a review, see Brannon, 2005) and early in development (even in newborns; e.g., Izard et al., 2009).

✉ Brian J. Scholl
brian.scholl@yale.edu

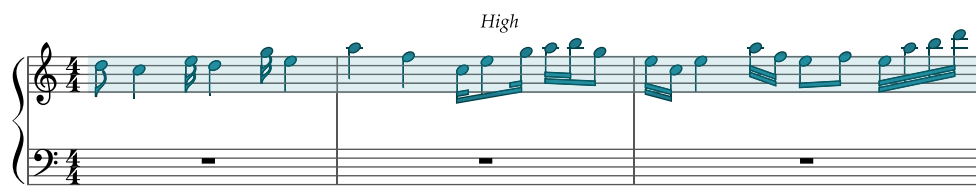
Joan Danielle K. Ongchoco
joan.ongchoco@yale.edu

¹ Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA

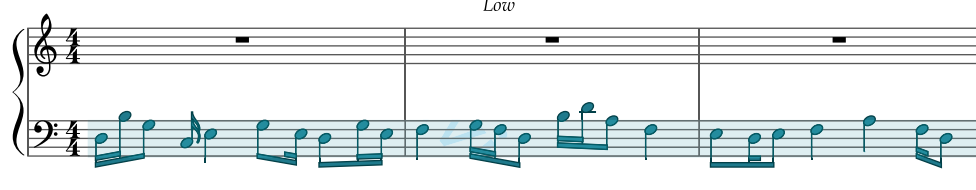
Sample Sequences

(a) Random

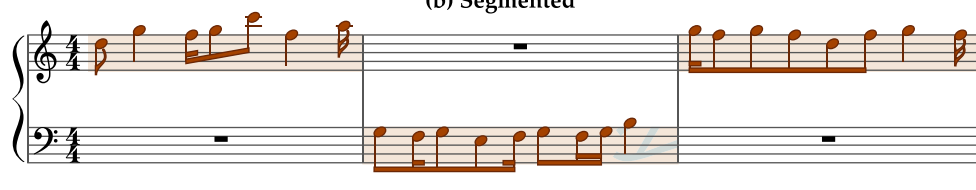
High



Low



(b) Segmented



(c) Gradual

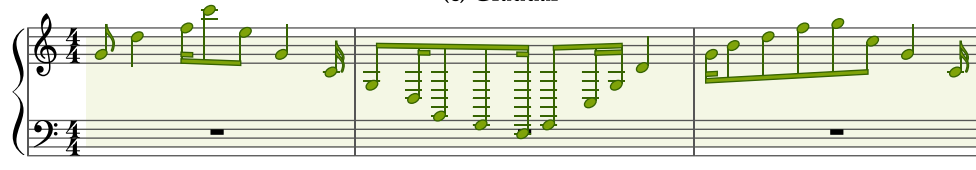


Fig. 2 Sample stimuli from the current experiments expressed in musical notation. (a) Random sequences, with notes drawn randomly from either a High or Low pitch-range. (b) Segmented sequences, with notes segmented into discrete musical phrases by sudden shifts in the

underlying range of the randomly selected pitches. (c) Gradual sequences, with gradually ascending and descending pitches, and with the highest and lowest notes within each measure matched to those from the Segmented sequences

The current project has a special focus on the units of enumeration in time, but it was directly inspired by previous work on the units of enumeration in space. Consider, for example, a seemingly simple task such as enumerating the number of discs in a display. If the display only contains discs, as in Fig. 1a, this task seems maximally natural. But how well can

you enumerate the number of discs in Fig. 1b – when some are connected into “dumbbells”? Here, a vast tradition of research in perceptual grouping entails that this “connectedness” cue is especially powerful at grouping the relevant pairs of discs into single (multi-disc) *objects* (e.g., Palmer & Rock, 1994; for a review, see Wagemans et al., 2012) – and this subsequent treatment of such dumbbells as single objects occurs to some degree irresistibly (Howe et al., 2012; Scholl et al., 2001). And indeed, if observers are asked to enumerate discs, they underestimate to a greater degree in displays such as Fig. 1b (Franconeri et al., 2009; He et al., 2009): they try to enumerate discs, but their minds can’t help enumerating dumbbells.

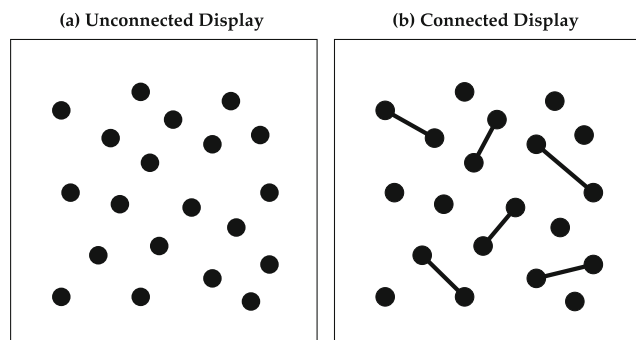


Fig. 1 Caricatures of displays from He et al. (2009). (a) A display with discs. (b) A display with the same number of discs, but with some pairs connected into dumbbells. When asked to enumerate the discs in such displays, observers underestimate to a greater degree with dumbbells (see also Franconeri et al., 2009)

Event segmentation

Previous work on the underlying units of mental enumeration (Franconeri et al., 2009; He et al., 2009; see also Yu et al., 2019; Zhao & Yu, 2016) has, to our knowledge, always involved static spatial displays. (Perhaps the only exception is work on “object persistence”; for a review, see Scholl, 2007.

But this work rarely stresses enumeration, and it involves set sizes of 1 or 2, rather than 20–40 as in the present work.)

But of course, our experience of the world is inherently dynamic – and just as spatial arrays are often represented in object-based terms, so too are temporal sequences often represented in terms of discrete (and temporally extended) *events*. In particular, we don't experience sequences in terms of an unstructured continuous flow of time, but as one (discrete) thing happening after another. And far from being arbitrary or idiosyncratic, people tend to agree on when discrete events start and end, even in naturalistic stimuli (e.g., Zacks et al., 2009). Moreover, event boundaries loom large for other mental processes such as attention and memory (e.g., Dubrow & Davachi, 2016; Hard et al., 2019; Heusser et al., 2018; Huff et al., 2012; Radvansky, 2012; Swallow et al., 2009).

The current study: What counts (in time)?

Enumeration must operate over discrete elements of some sort (by definition), but do we have full conscious, intentional control over what those units are? Can we enumerate whatever an experimenter asks us to enumerate, or must we in part enumerate what our visual system automatically treats as discrete events? Both possibilities seem plausible in different ways. On one hand, it seems natural to think that we can count whatever we choose to count – as in the infamous suggestion that if you want to know how many cows are in a field, just count the legs and divide by four. And this intuitive sense is also consistent with the empirical observation from the event-segmentation literature that observers can generally segment at whatever temporal scale they choose – varying their responses depending on whether they are asked to report the smallest or largest units of time (e.g., Newton, 1973; Zacks et al., 2001). On the other hand, it seems equally intuitive that you should be able to enumerate whatever elements in spatial arrays you choose (be they discs or dumbbells), but the empirical results reviewed above suggest that this is not so: even when you try to enumerate discs, your mind automatically seeks to enumerate dumbbells (Franconeri et al., 2009; He et al., 2009).

In the present study, we thus put this question to the empirical test, in a kind of “audio-temporal” analogue of the visuospatial “dumbbell” manipulation. Observers listened to sequences of musical notes, and simply had to report the number of individual notes that played on each trial. These notes were too numerous and were played too quickly and irregularly to support overt counting, so observers had to rely on an intuitive number sense. In some conditions, these notes were played in a single unstructured stream (as in Fig. 2a). But in other conditions, the notes were segmented into a few discrete musical phrases, by having their pitches grouped over time into different octaves (as in Fig.

2b). Critically, these musical phrases (like their dumbbell ancestors) were entirely task-irrelevant, and observers had no trouble whatsoever hearing the discrete notes. Could they choose to ignore the event segmentation cues, and simply focus on the notes? Or might they underestimate to a greater degree when the notes are segmented into discrete musical phrases?

Experiment 1: “How many notes played?”

In an initial experiment, observers estimated the number of notes that played in quick and irregular musical sequences, where the notes either were or were not segmented into a smaller number of musical phrases via shifts in the octaves of the tones (as in Fig. 2b).

Method

Participants Fifteen observers from the Yale and New Haven communities participated for credit or monetary payment. This preregistered sample size was determined via pilot experiments (exploring the same key contrast described below) before data collection began, and was fixed to be identical in each of the four experiments reported here.

Apparatus Stimuli were presented using custom software written in Python with the PsychoPy libraries (Peirce et al., 2019) and displayed on a monitor with a 60-Hz refresh rate. Observers sat in a dimly lit room without restraint approximately 60 cm from the display (with all visual extents reported below based on this approximate viewing distance). The functional part of the display subtended 34.87 x 28.21°. Observers listened to the musical sequences at a fixed volume (50% of the maximum possible volume) through headphones. Musical sequences were rendered using MuseScore software (exported to .ogg files that were then imported into PsychoPy).

Stimuli and conditions The speed with which notes were played was determined for each sequence by the number of “beats per minute” (BPM), which was randomly selected (separately for each sequence and each observer) to be between 108 and 156. Each sequence lasted for either 720/BPM, 960/BPM, 1,200/BPM, or 1,440/BPM s (such that a 720/BPM sequence with 120 BPM effectively lasted for 6 s), and each individual note in a given sequence had a randomly chosen duration of either 60/BPM, 30/BPM, or 15/BPM s. These three note durations were randomly intermixed (with no intervening silences) with the constraint that a new note had to play (regardless of the intervening notes) at exact intervals of 240/BPM s. (In musical terms, these same parameters can be summarized by noting that each sequence had between three and six 4/4-time-signature measures, with a random distribution of

“How many notes played?”

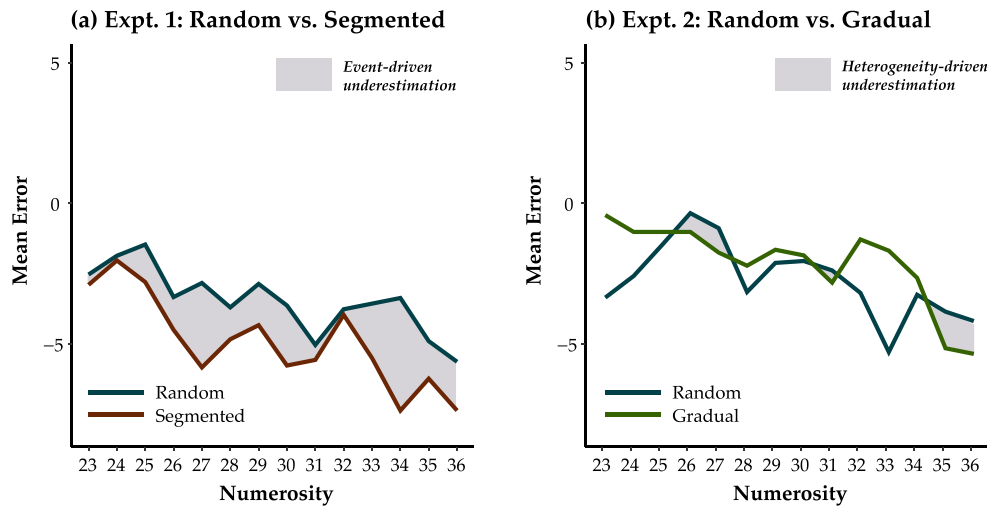


Fig. 3 Reported numerical estimates from (a) Experiment 1 (contrasting Random vs. Segmented sequences) and (b) Experiment 2 (contrasting Random vs. Gradual sequences). The vertical axis represents the mean errors (the actual numerosity subtracted from the reported estimate, such that positive numbers indicate overestimation, and negative numbers

quarter notes, eighth notes, and 16th notes – but with no rests – and with the constraint that a new note had to play on the downbeat of each measure.) A different set of four randomly generated sequences was pre-computed for each observer, for each of 14 numerosities – from 23 to 36 notes.

Each note was rendered via the “Piano” instrument setting in MuseScore, with the individual pitches depending on the condition. In Random sequences, each pitch was sampled randomly (differently for each sequence and each observer) from within either a High range (523–1,976 Hz) or a Low range (33–123 Hz), with an equal number of High- and Low-range sequences, and with the constraint that no two consecutive pitches could be identical.¹ In Segmented sequences, each pitch was sampled randomly (again differently for each sequence and each observer) from within a given pitch range, but that range alternated between the High and Low ranges every 240/BPM s. (In musical terms, these same parameters can be summarized by noting that Random pitches were randomly drawn from the notes of a C-major scale (i.e., from the white keys on a piano keyboard) within a two-octave range (High = from C5 to B6; Low = from C1 to B2), and that Segmented pitches were randomly drawn from two-octave ranges that alternated between High and Low every other measure. These conditions are depicted in musical notation in Fig. 2).

¹ The 14 possible High-range pitches were 523, 587, 659, 698, 784, 880, 988, 1,047, 1,175, 1,319, 1,397, 1,568, 1,760, and 1,976 Hz. And the 14 possible Low range pitches were 33, 37, 41, 44, 49, 55, 62, 65, 73, 82, 87, 98, 110, and 123 Hz.

indicate underestimation). The horizontal axis represents the actual number of notes that played on a given trial. The shaded regions represent event-driven underestimation (in Experiment 1) or heterogeneity-driven underestimation (or the lack thereof, in Experiment 2)

Procedure and design Each trial consisted of a single sequence of notes, with a visual display that was empty except for a centered black 3° x 3° sound icon on a white background. After the last note finished playing, this icon was replaced by the (centered, black, 0.5° height, Monaco font) response prompt “How many musical notes played in total?”, and observers typed in their response (between 1 and 99, with these digits appearing as they typed 2° below the response prompt). Observers then pressed a key to submit their response, after which their response turned purple for 500 ms, followed by a blank 500-ms interval before the next trial began.

Observers completed a single 30-note Random-sequence practice trial (the results from which were not recorded), followed by 56 experimental trials (14 numerosities (23–36) x 2 conditions (Random, Segmented) x 2 repetitions), presented in a different random order for each observer, with a short self-timed break after 28 trials.

Results

The error on each trial was computed by subtracting the actual number of notes in the musical sequence from the observer’s estimate of this value. (As such, negative errors indicate underestimation, and positive errors indicate overestimation – though as usual in such tasks with similar numerosities, observers always underestimated.) The resulting mean error rates for each of the conditions are depicted in Fig. 3a, broken down by each individual numerosity. Inspection of this figure reveals that observers underestimated more in Segmented sequences than in Random sequences — with the gray shading

between the two lines in Fig. 3a representing the magnitude of this event-driven underestimation. This difference was reliable both when comparing the average estimates ($t(14)=3.24$, $p=.006$, $d=0.84$) and when comparing the direction of this difference across the different numerosities (14/14; two-tailed binomial test, $p<.001$).

Discussion

These initial results are consistent with the possibility that observers experience the notes as segmented into a smaller number of discrete musical phrases, and that these phrases then serve as natural underlying units of enumeration, such that observers underestimate to a greater degree when such phrases exist compared to when a sequence is simply experienced as a single undifferentiated series of notes. (Of course, just as with the “dumbbells” in the spatial phenomena that inspired this project, we are not claiming that observers enumerate *only* the number of phrases – in which case they would produce much more radical underestimates. Instead, this effect seems to be probabilistic to some degree, with the central result being that segmentation into discrete multi-note events yields underestimation to some degree – and despite the observers’ explicit attempts to enumerate only the notes themselves.)

Experiment 2: Controlling for pitch heterogeneity

The central result of Experiment 1 was that observers underestimated to a greater degree in Segmented sequences, perhaps because (a) there were fewer musical phrases than individual notes in those sequences, (b) observers experienced the stimuli in terms of those musical phrases, and so (c) enumeration processes operated naturally over those experienced “units” of perception to some degree, yielding more underestimation relative to Random sequences. But another possibility is more mundane: perhaps this difference simply reflected the fact that the pitches in Segmented sequences were drawn from a wider range (spanning four octaves) than were the pitches in Random sequences (spanning two octaves) – consistent with the observation that heterogeneity influences enumeration (e.g., Marchant et al., 2013). To rule out this confound in the present experiment, we contrasted Random sequences with new Gradual control sequences, in which these pitch ranges were equated (to those in Segmented sequences) without introducing any segmentation into discrete musical phrases.

Method

This experiment was identical to Experiment 1, except as noted. Fifteen new observers participated, with this sample size chosen to match that of Experiment 1. Each individual Segmented sequence from Experiment 1 was transformed into a new Gradual sequence – with each note matched for duration (such that the underlying “rhythms” of these two sequences were identical), but with new pitches. In particular, the pitches of the notes at the middle of each 240/BPM s interval were set to either the highest pitch from that Segmented interval (if sampling from the High range) or the lowest pitch from that Segmented interval (if sampling from the Low range) — with the intervening pitches then ascending or descending in order, as depicted in Fig. 2c. (The intervening pitches were first evenly distributed between the highest and lowest pitches – to the nearest of the choices listed in Footnote 1 – and these values were then each randomly jittered by up to three steps, with the constraint that the sequence as a whole still had to gradually ascend or descend.)

Results and discussion

The mean error rates for each of the conditions are depicted in Fig. 3b, broken down by each individual numerosity. Inspection of this figure reveals very little selective underestimation driven by the Gradual condition over and above the Random condition (as depicted by the gray shading), in stark contrast to Experiment 1 and Fig. 3a. There was no reliable difference between Random and Gradual estimates ($t(14)=0.93$, $p=.366$, $d=0.24$) – and this null effect was reliably different from the significant degree of event-driven underestimation (i.e., from the Random/Segmented difference) in Experiment 1 ($t(28)=2.65$, $p=.013$, $d=0.97$). This confirms that the results from Experiment 1 did not merely reflect some new effect of pitch heterogeneity on underestimation.

Experiment 3: Controlling for duration (random versus segmented)

The task in Experiments 1 and 2 was to estimate the number of notes that played, but sequences with more notes naturally lasted longer on average – and so it is possible that observers could strategically base their responses not on enumeration but rather on a sense of each sequence’s duration. Although this wouldn’t predict the specific pattern of underestimation that we observed, we nevertheless wanted to ensure that the results reflected enumeration per se. Thus, this experiment replicated Experiment 1, but now with the underlying tempos manipulated such that all sequences (regardless of the number of notes) were well matched for duration.

Method

This experiment was identical to Experiment 1, except that Random and Segmented sequences were equated for duration by varying their tempos (between 80 and 168 BPM) such that each sequence's duration was between 9 and 10 s. Fifteen new observers participated, with this sample size chosen to match that of Experiments 1 and 2.

Results and discussion

The mean error rates for each of the conditions are depicted in Fig. 4a, broken down by each individual numerosity. Inspection of this figure reveals the same pattern that was observed in Experiment 1 and Fig. 3a (with the gray shading again representing the magnitude of this event-driven underestimation). The Random/Segmented difference was again reliable both when comparing the average estimates ($t(14)=2.73$, $p=.016$, $d=0.71$) and when comparing the direction of this difference across the different numerosities (12/14; two-tailed binomial test, $p=.013$). These results confirm that the event-driven underestimation cannot be due to some function of perceived duration.

Experiment 4: Controlling for duration (random versus gradual)

This experiment replicated the Random/Gradual contrast from Experiment 2, while controlling for duration in the same manner as in Experiment 3.

Method

This experiment was identical to Experiment 2, except that the sequence durations were determined as in Experiment 3. Fifteen new observers participated, with this sample size chosen to match that of Experiments 1, 2, and 3.

Results

The mean error rates for each of the conditions are depicted in Fig. 4b, broken down by each individual numerosity. Inspection of this figure reveals the same pattern that was observed in Experiment 2 and Fig. 3b, with no systematic difference in underestimation between the Random and Gradual sequences ($t(14)=0.60$, $p=.558$, $d=0.16$). Again (just as was true for the contrast between Experiments 1 and 2), this null effect was reliably different from the significant degree of event-driven underestimation (i.e., from the Random/Segmented difference) in Experiment 3 ($t(28)=2.43$, $p=.022$, $d=0.89$). This confirms that the results from Experiment 3 did not merely reflect some new effect of pitch heterogeneity on underestimation (while controlling for duration).

General discussion

A melody, practically by definition, is a series of individual notes. But we don't *experience* music as merely an unstructured series of notes. (If we did, that wouldn't be such an enjoyable and engaging experience!) Instead, the notes are

Controlling for Duration

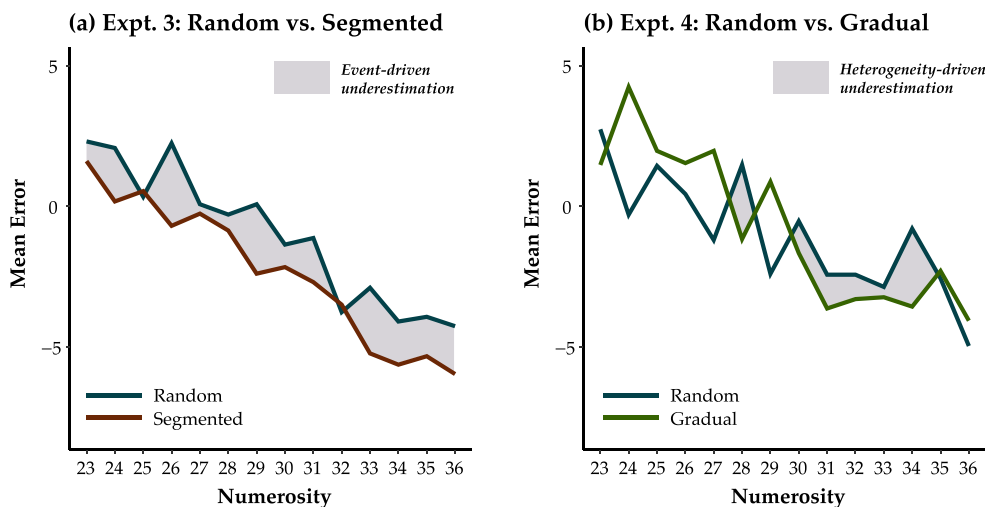


Fig. 4 Reported numerical estimates from (a) Experiment 3 (contrasting Random vs. Segmented sequences) and (b) Experiment 4 (contrasting Random vs. Gradual sequences). The vertical axis represents the mean errors (the actual numerosity subtracted from the reported estimate, such that positive numbers indicate overestimation, and negative numbers

indicate underestimation). The horizontal axis represents the actual number of notes that played on a given trial. The shaded regions represent event-driven underestimation (in Experiment 3) or heterogeneity-driven underestimation (or the lack thereof, in Experiment 4)

just lower-level building blocks of what we do experience, which are musical phrases (and motifs, and movements, etc.; for a review, see Bregman, 1990). Here, we asked how these experiences might irresistibly influence “what counts” when you’re trying to enumerate in time. In particular, observers were asked to enumerate the individual notes played in auditory sequences, but the higher-level events that they experienced (i.e., the musical phrases, as defined by sudden pitch-range changes) nevertheless influenced their performance – with greater underestimation in sequences that were segmented into a smaller number of phrases compared to sequences that had no such structure (even when matched for pitch range, pitch heterogeneity, and sequence duration). A natural interpretation of this result is that while observers attempted to dutifully enumerate notes, their minds were (at least to some modest degree, or with some probability) enumerating pseudo-musical phrases – i.e., those events that they actually experienced.

All of this is exactly analogous to previously observed results in the spatial domain. An object, practically by definition, is a collection of individual contours and edges. But we don’t *experience* visual scenes as merely an unstructured pattern of individual edges. Instead, the contours are just lower-level building blocks of what we do experience, which are objects (and parts, and groups, etc.; for a review, see Wagemans et al., 2012). And so when observers are asked to enumerate the individual shapes in a display with more complex objects (e.g., enumerating the discs in a display where many are connected into dumbbells) they underestimate to a greater degree – as if they attempted to dutifully enumerate discs, but their minds were (at least partly) enumerating dumbbells – i.e., those objects that they actually *saw* (Franconeri et al., 2009; He et al., 2009).

(Of course, connectedness is not the only grouping cue that contributes to objecthood, and so we might imagine also obtaining this “object-driven underestimation” with other cues such as continuity or closure; Feldman, 2007; Marino & Scholl, 2005. Similarly here, pitch-range shifts are presumably not the only cue that contributes to event segmentation, and so we predict that “event-driven underestimation” could also be driven by sudden changes in other factors such as timbre – such that the sudden (vs. gradual) shift from a trumpet to a xylophone might also lead to robust event segmentation and the experience of a new musical phrase.)

The present results are thus an exact “audio-temporal” analogue of the previously observed visuospatial phenomenon of underestimation with “dumbbells.” In both cases, what these results show is that we don’t have full conscious, intentional control over “what counts”; we can’t simply enumerate whatever clearly discernable elements an experimenter asks us to enumerate. Instead, what is enumerated in the first place is

to some degree determined by how scenes (be they visual or auditory, spatial or temporal) are automatically segmented into the particular discrete individual objects or events that we naturally experience.

Author Note For helpful conversations and/or comments on earlier drafts, we thank Gabriel Radvansky, Yaoda Xu, and the members of the Yale Perception & Cognition Laboratory. This project was funded by ONR MURI #N00014-16-1-2007 awarded to BJS. The preregistered methods and analyses for each experiment can be viewed at <http://aspredicted.org/blind.php?x=qg6se2> and <http://aspredicted.org/blind.php?x=ph9jv6>.

Author Contributions J. D. K. Ongchoco and B. J. Scholl designed the research and wrote the manuscript. J. D. K. Ongchoco conducted the experiments and analyzed the data with input from B. J. Scholl.

References

- Brannon, E. M. (2005). What animals know about numbers. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 85–107). Psychology Press.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford University Press.
- DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning & Memory*, *134*, 107–114.
- Feldman, J. (2007). Formation of visual “objects” in the early computation of spatial relations. *Perception & Psychophysics*, *69*, 816–827.
- Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*, 1–13.
- Hard, B. M., Meyer, M., & Baldwin, D. (2019). Attention reorganizes as structure is detected in dynamic action. *Memory & Cognition*, *47*, 17–32.
- He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review*, *16*, 509–517.
- Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *44*, 1075–1090.
- Howe, P. D., Incledon, N. C., & Little, D. R. (2012). Can attention be confined to just part of a moving object? Revisiting target-distractor merging in multiple object tracking. *PloS One*, *7*, e41491.
- Huff, M., Papenmeier, F., & Zacks, J. M. (2012). Visual target detection is impaired at event boundaries. *Visual Cognition*, *20*, 848–864.
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*, 10382–10385.
- Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, *142*, 245–250.
- Marino, A. C., & Scholl, B. J. (2005). The role of closure in defining the “objects” of object-based attention. *Perception & Psychophysics*, *67*, 1140–1149.
- Newton, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality & Social Psychology*, *28*, 28–38.

- Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*, 29–55.
- Peirce, J. W., Gray, J., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.
- Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, *21*, 269–272.
- Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.
- Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind & Language*, *22*, 563–591.
- Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159–177.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*, 236–257.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*, 1172–1217.
- Yu, D., Xiao, X., Bemis, D., & Franconeri, S. (2019). Similarity grouping as feature-based selection. *Psychological Science*, *30*, 376–385.
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*, 307–327.
- Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*, 29–58.
- Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, *146*, 217–222.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.