Supplemental Material: The complexity-aesthetics relationship for musical rhythm is more fixed than flexible: Evidence from children and expert dancers

Daniel J Cameron[1], Nicole Caldarone[1], Maya Psaris[1], Chantal Carrillo[1], Laurel J Trainor[1,2,3]

[1]Dept. of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Canada
[2]McMaster Institute for Music and the Mind, McMaster University, Hamilton, Canada
[3]Rotman Research Institute, Baycrest Hospital, Toronto, Canada

Experiment 1:

Does the Inverted-U Rely on Musically Implausible High Syncopation Rhythms?

To test whether the musically implausible High syncopation drum breaks (those which were experimenter-composed) had a significant effect on the observed inverted-U relationship between syncopation and groove, we reanalyzed dancers and nondancers quadratic fits with and without those stimulus items, and compared the results quadratic and linear goodness-of-fit (adjusted r^2) for all participants combined in a 2x2 repeated-measures ANOVA for the factors Stimulus Set (full vs. omitting musically implausible high rhythms) and Fit Type (quadratic vs. linear). We found main effects of Stimulus Set (goodness-of-fit, or r^2, was greater for the full stimulus set, $F(1, 380) = 28.60$, $p = <.001$), Fit Type (quadratic fits had greater goodness-of-fit than linear fits, $F(1, 380) = 135.28$, $p = <.001$), and an interaction between the two factors ($F(1, 380) = 56.47$, $p = <.001$). Driving the interaction, the gain in goodness-of-fit by quadratic compared to linear fits was greater for the full stimulus set than the stimulus set that omitted the musically-implausible High syncopation items (see Fig. S1). Follow-up paired *t*-tests indicate that quadratic fits had greater goodness-of-fit than linear fits for both the full stimulus set ($t(188) = 10.97$, $p < .001$) and the stimulus set omitting the musically implausible High syncopation stimuli ($t(188) = 4.31$, $p < .001$). This result indicates that these particular stimulus items, which are very high in syncopation but also musically implausible, are not required to observe the inverted-U syncopation-groove relationship, although they exaggerate it.

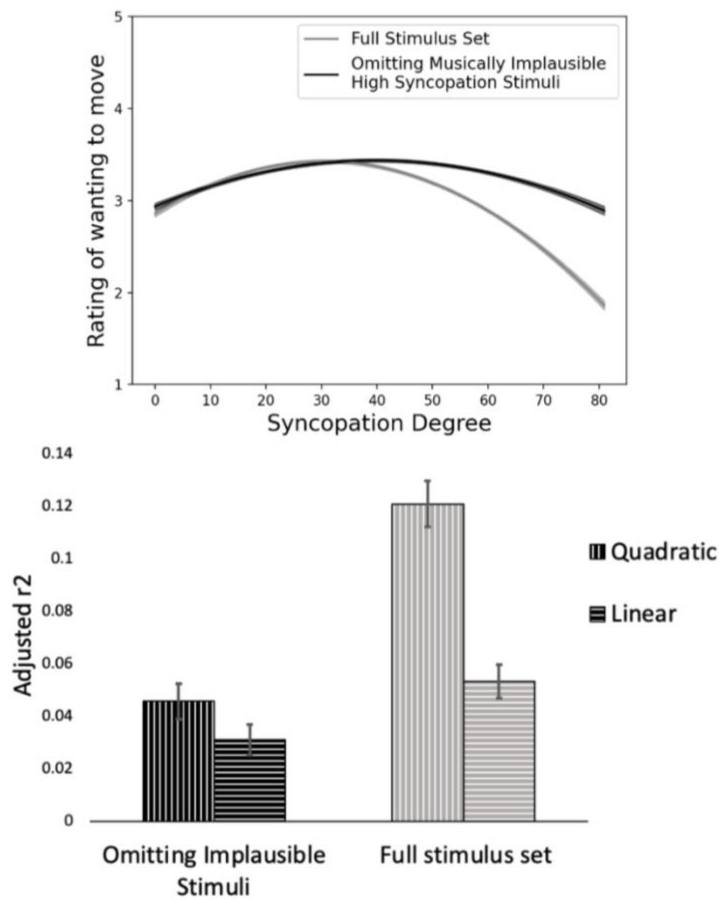

Figure S1: The inverted-U syncopation-groove relationship does not depend on musically implausible high-syncopation stimulus items (although it is exaggerated by them). Top: quadratic fits (bold lines) and 95% confidence intervals (corresponding shaded areas) for average groove ratings with respect to syncopation. Bottom: Mean goodness-of-fit (adjusted r^2) for quadratic and linear fits for groove ratings when including vs. excluding the musically implausible high syncopation stimuli.

Correlations between Groove and Pleasure in Nondancers

We tested for association between the urge to move and pleasure in two different ways. First, we calculated the Pearson correlation coefficient between these two ratings across all 50 stimulus rhythms for each participant (nondancers only). The average within-subject Groove-Pleasure correlation was $r = .357$ (SD = .253), $p < .0001$. We also calculated the Pearson correlation coefficient of the Groove and Pleasure ratings for the 50 rhythms (averaged across participants for each rhythm). Average Groove and Pleasure ratings were highly correlated across the 50 rhythms ($r = .98$, $p < .0001$).

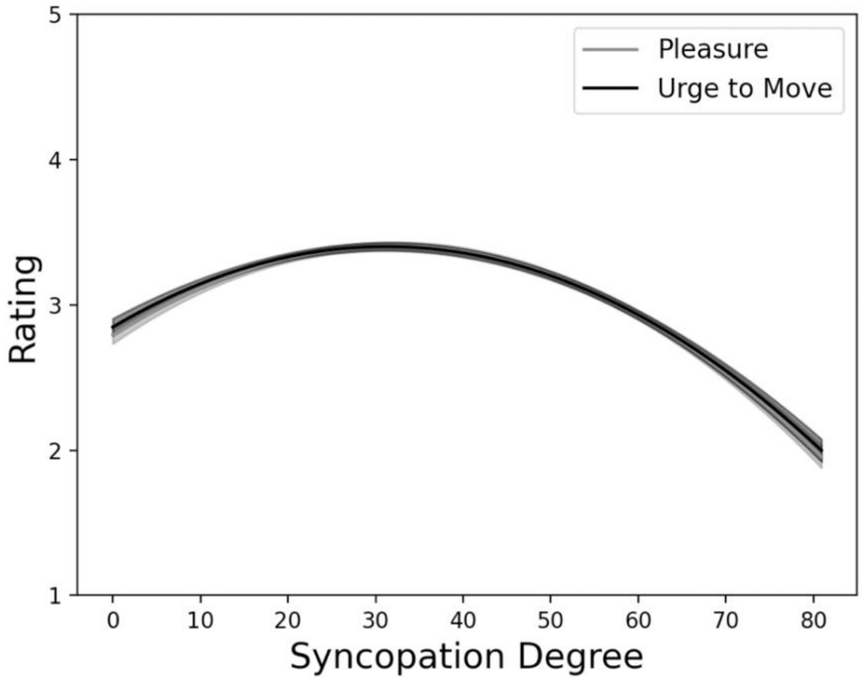

Figure S2. The two elements of groove, the urge to move and pleasure, are correlated and appear to have very similar quadratic fits. Bold lines (in dark and light grey) indicate mean fits for ratings of the urge to move and of pleasure. Shaded areas around each bold line indicate 95% confidence intervals.

Comparing the Urge to Move with the Urge to Dance in Dancers

To compare whether groove, typically considered a general urge to move along to music, has the same relatinship to syncopation as the urge to dance, we compared Dancers' ratings and aspects of their quadratic fits for the ratings questions of how much rhythms make the listener want to move in a general way versus want to dance in the specific style with which they have expertise using two-tailed independent samples *t*-tests (see Fig. S3). Mean ratings were higher for the general urge to move compared to the urge to dance ($t(41) = 5.06$, $p < .001$). The optimal level of syncopation trended to be higher for urge to move versus urge to dance ($t(41) = 1.77$, $p = .084$). Quadratic trends were stronger for the urge to move compared to the urge to dance ($t(41) = 3.50$, $p = .001$), as was goodness-of-fit ($t(41) = 2.27$, $p = .028$). Together, these results show that the general urge to move is distinct from a specific urge to dance, and that a general urge to move has a stronger and more robust inverted-U relationship with syncopation and possibly with a higher optimal syncopation.

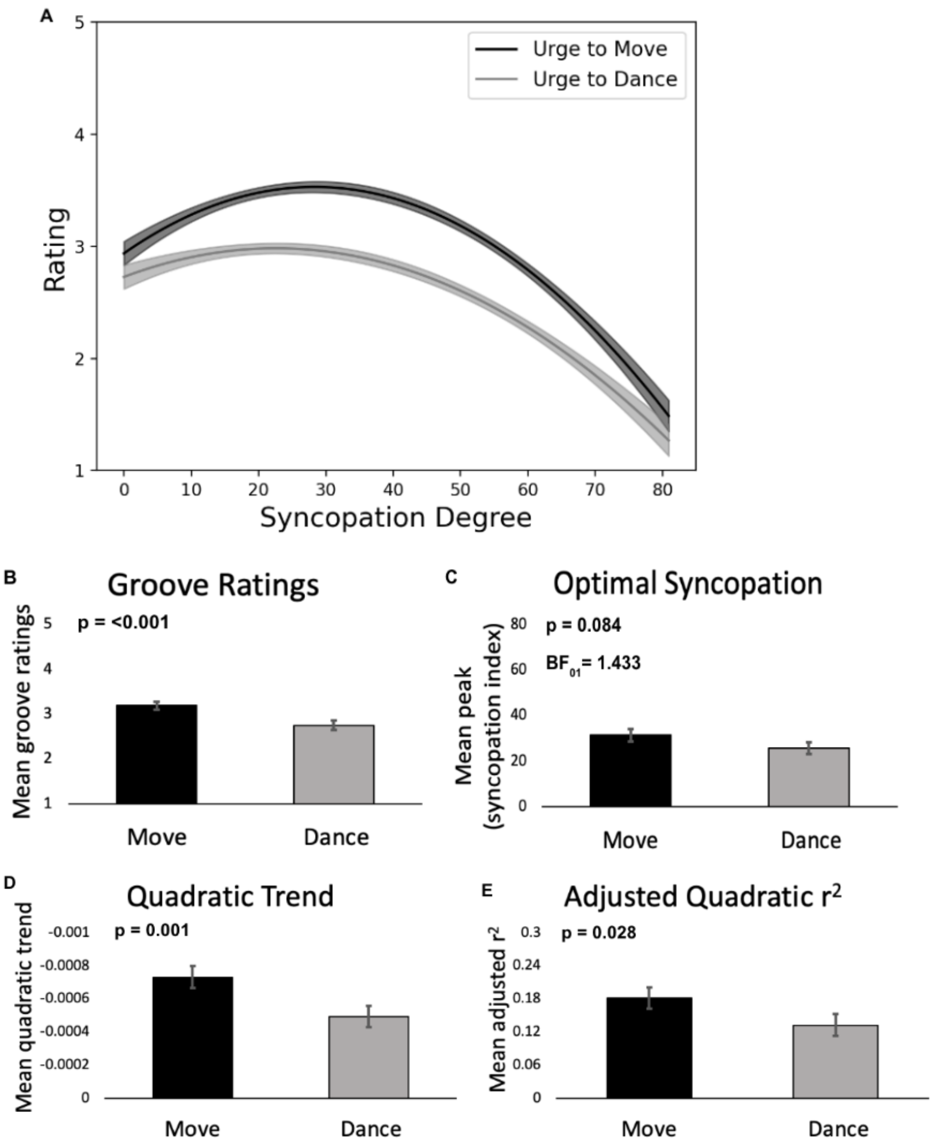

Figure S3: Groove (the urge to move to rhythm) is distinct from the urge to dance in one's dance style of expertise. A) Dancers' mean quadratic fits and 95% confidence intervals for the two types of ratings: the general urge to move, and the specific urge to dance in one's style of expertise. B-E: Means (error bars indicate ±1 SEM) for B) mean groove ratings, C) optimal syncopation (peak of the quadratic fit), D) quadratic trend (more negative is a sharper inverted-U shape), and E) adjusted r^2 for quadratic fits.

Correlations Between Syncopation-Groove Measures, Dance Sophistication, and Personality

The full set of Spearman correlation coefficients and corresponding FDR-corrected *p* values are found in Table S1, below.

Move avg ratings															
Move Peak	ρ = -0.023 pFDR = 0.952														
Move adjusted r2	ρ = -0.022 pFDR = 0.952	**ρ = -0.240 pFDR = 0.006**													
Move quadratic trend	ρ = -0.034 pFDR = 0.899	ρ = 0.026 pFDR = 0.933	ρ = -0.695 pFDR = 0												
Active Dance	ρ = 0.073 pFDR = 0.570	ρ = -0.087 pFDR = 0.471	**ρ = 0.202 pFDR = 0.026**	**ρ = -0.232 pFDR = 0.008**											
Passive Dance	ρ = 0.052 pFDR = 0.747	ρ = -0.012 pFDR = 1	ρ = 0.177 pFDR = 0.060	ρ = -0.154 pFDR = 0.114	ρ = 0.579 pFDR = 0										
Body Awareness	ρ = 0.107 pFDR = 0.340	ρ = -0.119 pFDR = 0.284	ρ = 0.063 pFDR = 0.654	ρ = -0.154 pFDR = 0.114	ρ = 0.552 pFDR = 0	**ρ = 0.420 pFDR = > 0.001**									
Social Dance	ρ = -0.004 pFDR = 1	ρ = -0.041 pFDR = 0.838	ρ = -0.045 pFDR = 0.811	ρ = -0.046 pFDR = 0.813	**ρ = 0.354 pFDR = > 0.001**	ρ = 0.094 pFDR = 0.418	ρ = 0.142 pFDR = 0.151								
Urge to Dance	ρ = 0.140 pFDR = 0.154	ρ = -0.022 pFDR = 0.952	**ρ = 0.251 pFDR = 0.004**	**ρ = -0.306 pFDR = > 0.001**	ρ = 0.754 pFDR = 0	**ρ = 0.443 pFDR = > 0.001**	**ρ = 0.333 pFDR = > 0.0001**	**ρ = 0.203 pFDR = 0.026**							
Dance Training	ρ = 0.019 pFDR = 0.965	ρ = -0.080 pFDR = 0.527	**ρ = 0.202 pFDR = 0.026**	ρ = -0.156 pFDR = 0.114	ρ = 0.795 pFDR = 0	**ρ = 0.502 pFDR = > 0.001**	**ρ = 0.290 pFDR = > 0.0001**	ρ = -0.020 pFDR = 0.965	**ρ = 0.425 pFDR = > 0.001**						
Extraversion	ρ = 0.052 pFDR = 0.747	ρ = 0.151 pFDR = 0.122	ρ = -0.016 pFDR = 0.974	ρ = -0.001 pFDR = 1	ρ = 0.117 pFDR = 0.286	ρ = 0.093 pFDR = 0.422	ρ = -0.029 pFDR = 0.912	ρ = 0.087 pFDR = 0.471	ρ = 0.117 pFDR = 0.286	ρ = 0.100 pFDR = 0.365					
Agreeableness	ρ = -0.113 pFDR = 0.297	ρ = 0.072 pFDR = 0.583	ρ = 0.010 pFDR = 1	ρ = 0.020 pFDR = 1	**ρ = 0.196 pFDR = 0.032**	**ρ = 0.204 pFDR = 0.026**	ρ = 0.158 pFDR = 0.110	ρ = 0.015 pFDR = 0.981	**ρ = 0.199 pFDR = 0.029**	ρ = 0.142 pFDR = 0.151	ρ = 0.117 pFDR = 0.286				
Conscientiousness	ρ = 0.065 pFDR = 0.637	ρ = 0.075 pFDR = 0.560	ρ = 0.060 pFDR = 0.680	ρ = -0.009 pFDR = 1	ρ = 0.055 pFDR = 0.740	ρ = 0.104 pFDR = 0.355	ρ = 0.035 pFDR = 0.899	ρ = 0.044 pFDR = 0.814	ρ = 0.068 pFDR = 0.612	ρ = 0.028 pFDR = 0.913	ρ = 0.153 pFDR = 0.114	ρ = 0.165 pFDR = 0.093			
Emotional Stability	ρ = -0.052 pFDR = 0.747	ρ = 0.032 pFDR = 0.911	ρ = -0.006 pFDR = 1	ρ = 0.031 pFDR = 0.913	ρ = 0.106 pFDR = 0.341	ρ = 0.102 pFDR = 0.364	ρ = 0.109 pFDR = 0.332	ρ = 0.032 pFDR = 0.911	ρ = 0.083 pFDR = 0.497	ρ = 0.067 pFDR = 0.621	**ρ = 0.307 pFDR = > 0.001**	**ρ = 0.193 pFDR = 0.035**	**ρ = 0.184 pFDR = 0.048**		
Openess	ρ = 0.003 pFDR = 1	ρ = 0.115 pFDR = 0.291	ρ = 0.017 pFDR = 0.974	ρ = -0.021 pFDR = 0.952	ρ = 0.005 pFDR = 1	ρ = 0.079 pFDR = 0.536	ρ = 0.077 pFDR = 0.549	ρ = 0.048 pFDR = 0.786	ρ = -0.035 pFDR = 0.899	ρ = 0.030 pFDR = 0.913	ρ = 0.161 pFDR = 0.101	ρ = 0.100 pFDR = 0.365	**ρ = 0.236 pFDR = 0.007**	ρ = 0.148 pFDR = 0.132	
	Move avg ratings	**Move Peak**	**Move adjusted r2**	**Move quadratic trend**	**Active Dance**	**Passive Dance**	**Body Awareness**	**Social Dance**	**Urge to Dance**	**Dance Training**	**Extraversion**	**Agreeableness**	**Conscientiousness**	**Emotional Stability**	**Openess**

Table S1. Spearman correlation coefficients and *p* values corrected using the false discovery rate (FDR).

Comparing the Syncopation-Groove Relationship in Musicians and Nonmusicians

We compared the syncopation-groove relationships between musicians (≥ 8 years of training; n = 25) and nonmusicians (< 4 years of training; n = 80) from the sample of dancers and nondancers. We compared four measures of groove and its quadratic relationship to syncopation using one-tailed independent samples *t* tests (see Fig. S4). There was a statistical trend for musicians to have higher mean groove ratings than nonmusicians' ratings ($t(103) = 1.48$, $p = 0.071$). Optimal syncopation did not differ between groups ($p > .05$). There was a statistical trend for musicians to have stronger quadratic trends than nonmusicians ($t(103) = -1.50$, $p = 0.068$). Goodness-of-fit did not differ between groups ($p > .05$).

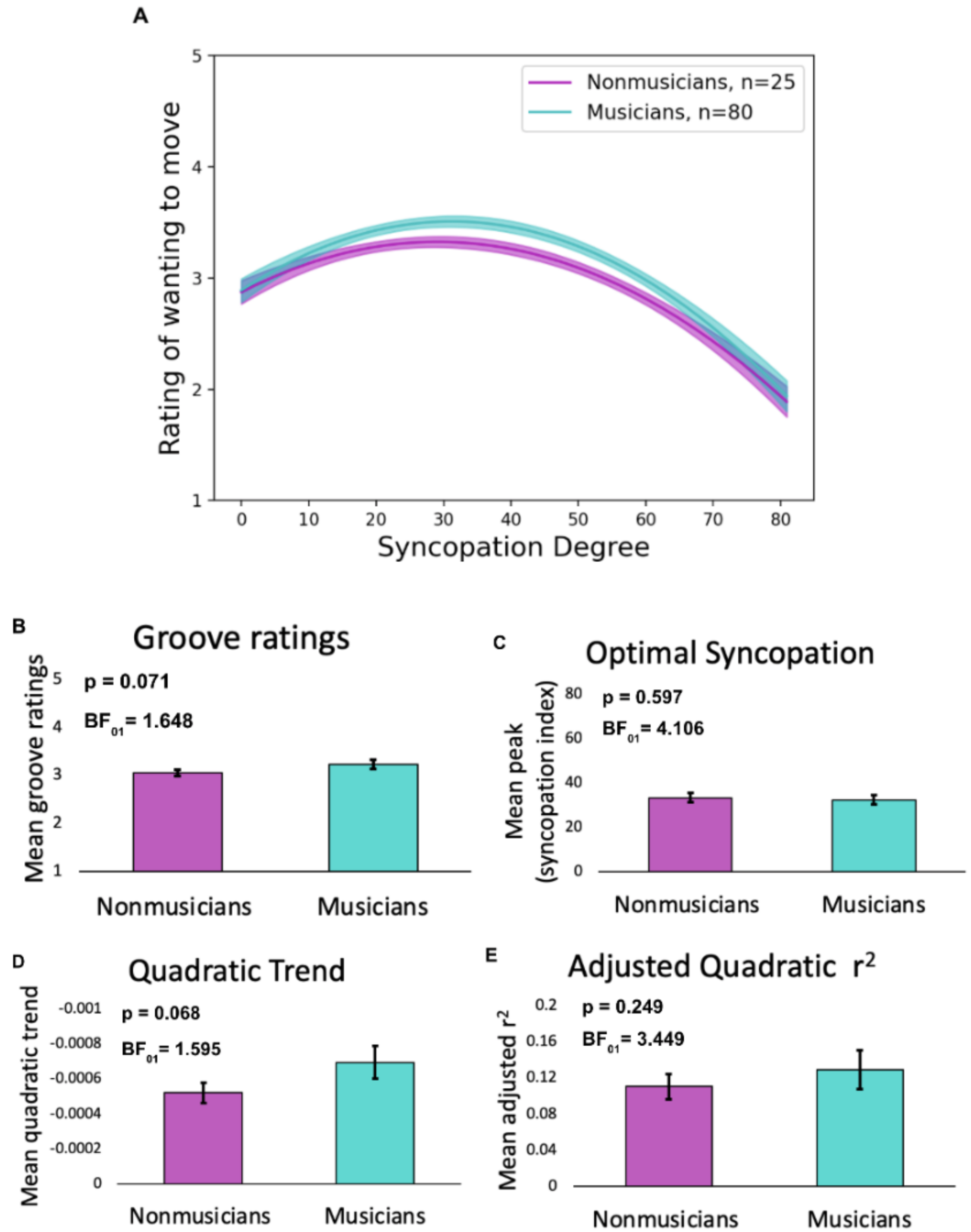

Figure S4: A) Musicians' and Nonmusicians' mean quadratic fits and 95% confidence intervals. B-E: Means (error bars indicate ±1 SEM) for B) mean groove ratings, C) optimal syncopation (peak of the quadratic fit), D) quadratic trend (more negative is a sharper inverted-U shape), and E) adjusted r^2 for quadratic fits.

Testing for Group Differences in Linear Fits

To test whether the linear relationship between syncopation and groove differed between groups, we used independent samples *t* tests. We tested for group differences in the slope and the goodness-of-fit (adjusted r^2) from the linear fits from each participant's groove ratings across the range of syncopation. We compared slopes and goodness-of-fit between Dancers and Nondancers, Ballet and Hip Hop dancers, and Musicians and Nonmusicians. Slopes were more negative for Dancers (r = -.014) than Nondancers (r = -.008) (t(65) = 2.74, p = .008), and linear adjusted r^2 was higher for Dancers (r^2 = .085) than Nondancers (r^2 = .038) (t(65) = 2.46, p = .017). Neither slope nor linear goodness-of-fit differed (p > .05) between Ballet and Hip Hop dancers or between Musicians and Nonmusicians. Previous work reported that musicians had a stronger negative linear relationship between syncopation and groove than nonmusicians (Matthews et al., 2022). Our results differ from that finding in that the nonexperts (nondancers in this case) had more negative linear syncopation-groove relationship than experts, and we found no difference between Musicians and Nonmusicians.

Experiments 2a and 2b: Children's Perception and Behaviour

Complete Instructions for Children's Forced-Choice Task

"We're going to have a birthday party and we want everyone to dance! We need you to help choose who will play the drums at the party, so that we find the best drumming for everyone to dance along to. Some animals are going to play drums for you, and you need to help decide which animals do the best drumming to dance to! We need you to listen very carefully to all the animals drumming and decide which are best for dancing. You will hear two animals in each trial. The two animals will look the same but they do different drumming. One of them is really good at drumming to make kids dance, and one is not good at drumming to make kids dance. YOU need to listen to both animals and pick which drumming is better for dancing."

Syncopation and Enjoyment in Children

To test whether children's enjoyment differed between levels of syncopation, we compared differences in response proportions for each of the three syncopation-comparison conditions from Experiment 2b. There were no statistically significant differences in enjoyment between syncopation levels (p > .05; see Fig. S5).

Differences in response proportions between levels of syncopation did not correlate with age (p > .05).

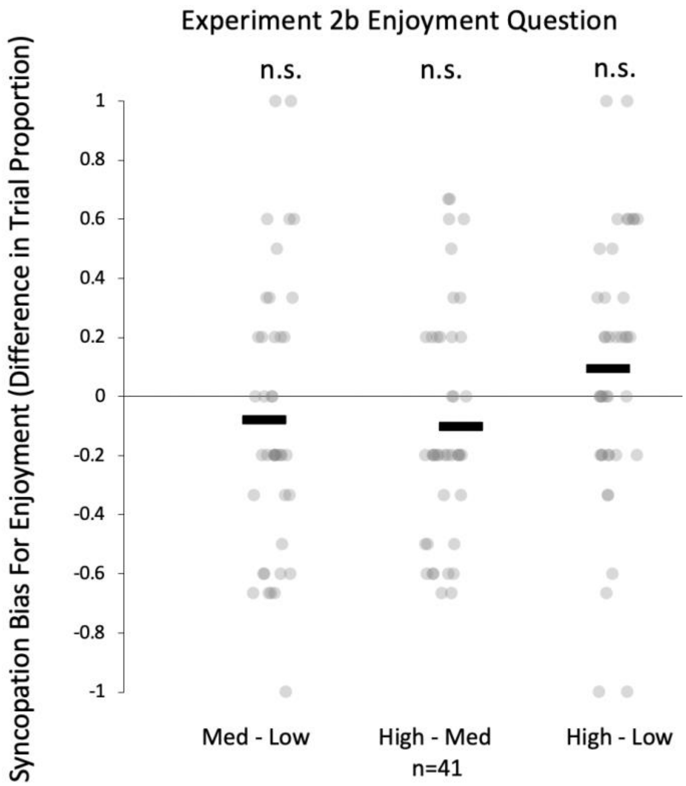

Figure S5: 3- to 6-year-old children did not show statistically significant enjoyment biases for syncopation. Grey circles indicate individual participants and horizontal black bars indicate means.

Experiment 2a Dance Task Results from Alternate Rating Questions

Raters assessed children's movements in the Dance task by watching videos and rating each video on three additional questions besides the primary rating question (results from which are reported in the main text) on scales from 1 to 7: "How fast were their movements?", "How big were their movements?", and "About how much of the trial were they dancing for?". For these three rating questions, we compared mean ratings for each question in each syncopation-comparison condition using one-tailed (for Low vs. Medium and Medium vs. High) and two-tailed (for Low vs. High) paired *t*-tests. All results for these questions were not significant ($p >$ .05).

Raters also rated how much space each child had to dance in and we compared these ratings across the three syncopation-comparison groups in a 1x3 ANOVA which indicated statistically significant differences between groups ($F(2,100) = 3.79$, *p*, = .026). Post-hoc *t*-tests indicated that there was no difference between the Low & Medium group and the Medium & High group ($p > .05$), but that the Low & Medium and Medium & High groups both had higher ratings of the size of the dance space participants were in compared to the Low & High group ($t(68) = 2.56$, $p = .013$; $t(64) = 2.15$, $p = .036$, respectively). To test whether these group differences were likely to have had an effect on the group differences in overall movement ratings (reported in the main text) we test for association between rated space and overall movement ratings (averaged across syncopation conditions) using Spearman correlation. No statistically significant correlations were found in any of the syncopation-comparison groups or when combining all groups (n = 103). Therefore, we believe it is unlikely that the amount of space individual participants had to dance in affected their movement and the reported group differences.

Testing for Effects of Event Density in Experiment 2a and 2b

To test whether event density (the number of notes in a stimulus rhythm) was associated with the urge to move, we tested for Pearson correlation between the event density of individual stimulus rhythms and their proportion of being chosen as eliciting more groove (adults) or better for dancing (children) in Experiments 2a and 2b. There were no statistically significant correlations between event density and response proportion for adults or children in any syncopation-comparison condition ($p > .05$).

Testing for Effects of Specific Stimulus Rhythms

To test whether individual stimulus items had greater effects than others and may have driven the condition effects (e.g., effects of Medium over Low syncopation) we compared the mean response proportions and movement ratings of the four individual rhythms within each syncopation condition (Low, Medium, and High) using 1x4 ANOVA, separately for children and adults, correcting for multiple comparisons using the false discovery rate (FDR). In all cases (all pairings of all syncopation conditions), no significant differences were reported ($p_{FDR} > .05$).