



# Role of the striatum in incidental learning of sound categories

Sung-Joo Lim<sup>a,b,c,1</sup>, Julie A. Fiez<sup>b,c,d,e</sup>, and Lori L. Holt<sup>a,b,c,d</sup>

<sup>a</sup>Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213; <sup>b</sup>Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213; <sup>c</sup>Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA 15213; <sup>d</sup>Center for Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260; and <sup>e</sup>Department of Psychology, University of Pittsburgh, PA 15260

Edited by Richard N. Aslin, Haskins Laboratories, New Haven, CT, and approved January 18, 2019 (received for review July 17, 2018)

Humans are born as “universal listeners” without a bias toward any particular language. However, over the first year of life, infants’ perception is shaped by learning native speech categories. Acoustically different sounds—such as the same word produced by different speakers—come to be treated as functionally equivalent. In natural environments, these categories often emerge incidentally without overt categorization or explicit feedback. However, the neural substrates of category learning have been investigated almost exclusively using overt categorization tasks with explicit feedback about categorization decisions. Here, we examined whether the striatum, previously implicated in category learning, contributes to incidental acquisition of sound categories. In the fMRI scanner, participants played a videogame in which sound category exemplars aligned with game actions and events, allowing sound categories to incidentally support successful game play. An experimental group heard nonspeech sound exemplars drawn from coherent category spaces, whereas a control group heard acoustically similar sounds drawn from a less structured space. Although the groups exhibited similar in-game performance, generalization of sound category learning and activation of the posterior striatum were significantly greater in the experimental than control group. Moreover, the experimental group showed brain–behavior relationships related to the generalization of all categories, while in the control group these relationships were restricted to the categories with structured sound distributions. Together, these results demonstrate that the striatum, through its interactions with the left superior temporal sulcus, contributes to incidental acquisition of sound category representations emerging from naturalistic learning environments.

auditory categorization | incidental category learning | distributional regularity | fMRI | corticostriatal systems

Categorization is a powerful cognitive process that supports the differentiation and interpretation of objects and events in the environment. Learning categories supports generalization of knowledge to new, unfamiliar instances by allowing organisms to come to treat physically and perceptually distinct objects—that nonetheless share deeper statistical structure—as functionally equivalent. Because categorization is so central to cognition, it is important to understand the processes and neural substrates that support human category learning. One candidate neural substrate is the striatum, the input structure of the basal ganglia (1). Here, our goal is to investigate the role of the striatum in auditory category learning, using a videogame task in which category learning occurs incidentally, without explicit training. This creates learning conditions similar to those experienced during speech learning, thereby building a bridge between one of the brain’s central learning systems and one of our most important cognitive abilities.

Knowledge about the neural mechanisms that support category learning has come largely from studies of visual category learning under explicit training. In these tasks, participants actively attempt to assign exemplars to the correct category through a motor response, and use explicit feedback about category decisions to guide learning. Visual category learning under explicit training conditions

is thought to involve the striatum and associated corticostriatal loops by which the basal ganglia interact with visual and motor cortices, especially when the learning involves complex perceptual categories that are difficult to describe verbally (2–6).

A much smaller body of work has examined the neurobiological basis of auditory category learning. To date, most of this work has focused on identifying changes in the cerebral cortical response to sounds after learning (7–11). Thus, while we have gained substantial knowledge about the cortical locus of learning-related change in auditory and speech categorization, little is known about the learning mechanisms and associated brain systems that drive these changes in auditory categorization ability.

An exception comes from studies of categorization training among adults learning difficult-to-acquire nonnative speech categories. These studies show that explicit training with feedback can produce significantly larger learning gains than those yielded from passive exposure (12–14). Critically, this behavioral dissociation is reflected in the striatum during learning (12); training with explicit feedback results in striatal activation, whereas training with no feedback does not. Therefore, at least under conditions of explicit training with feedback, the striatum is a likely contributor to speech category learning (12, 15, 16).

Thus far, there is clear evidence that explicit feedback-based categorization training can involve the striatum. One important limitation, however, is that we do not yet know whether the striatum is engaged in incidental category learning under more

## Significance

Humans are born as “universal listeners.” However, over the first year, infants’ perception is shaped by native speech categories. How do these categories naturally emerge without explicit training or overt feedback? Using fMRI, we examined the neural basis of incidental sound category learning as participants played a videogame in which sound category exemplars had functional utility in guiding videogame success. Even without explicit categorization of the sounds, participants learned functionally relevant sound categories that generalized to novel exemplars when exemplars had an organized distributional structure. Critically, the striatum was engaged and functionally connected to the auditory cortex during game play, and this activity and connectivity predicted the learning outcome. These findings elucidate the neural mechanism by which humans incidentally learn “real-world” categories.

Author contributions: S.-J.L., J.A.F., and L.L.H. designed research; S.-J.L. performed research; S.-J.L. analyzed data; and S.-J.L., J.A.F., and L.L.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence should be addressed. Email: sungj.m.lim@gmail.com.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811992116/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811992116/-DCSupplemental).

Published online February 19, 2019.

ecologically valid learning conditions, such as active engagement in a task that involves multiple forms of sensory input, some of which may be coherently structured but not explicitly categorized. The ability to learn in such environments is particularly crucial for understanding auditory category learning, because humans readily acquire native speech categories without instruction or explicit feedback.

Moreover, there remain important unanswered questions about how incidental category learning may be influenced by the distributional structure of category input. Distributional regularities in the input impact learning strategies and outcomes in explicit training paradigms (e.g., refs. 17 and 18). Likewise, the striatum has been implicated in procedural learning of fixed stimulus–response associations across simple stimulus classes via rewarding outcomes, even without explicit awareness (e.g., refs. 1, 19, and 20). However, in the former case, explicit training does not model the challenge of speech learning. In the latter, the simplicity of the mapping from stimulus to reward does not adequately model the challenge of mapping from complex but distributionally structured speech category input to reward.

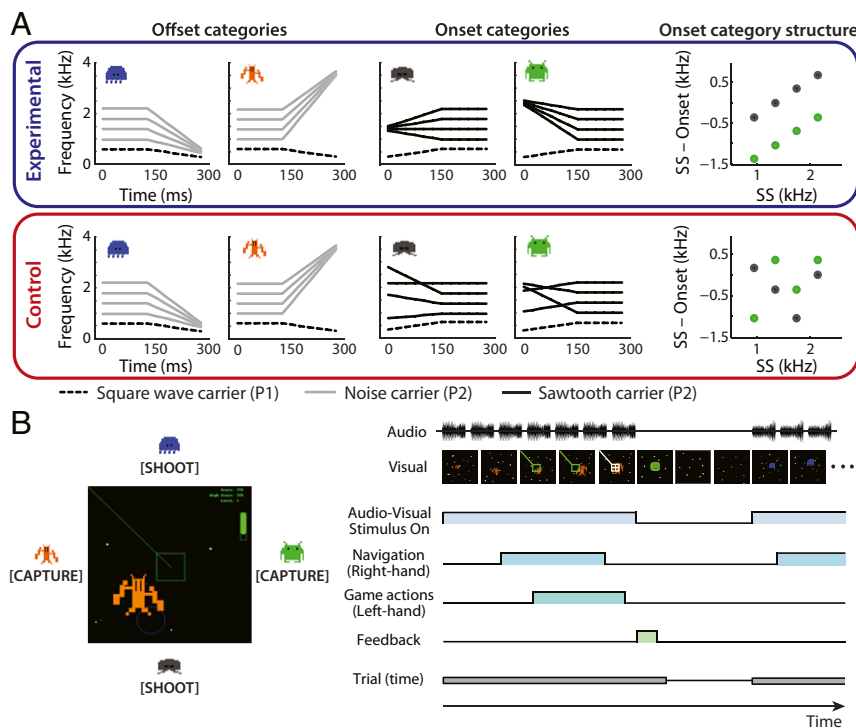
This leads us to examine whether the striatum is involved in learning the distributional structure of auditory categories in a rich, multimodal environment without overt categorization. We take a novel approach in which participants play a videogame during an fMRI scan. In the course of game play, participants learn auditory categories incidentally, through the categories' utility in supporting successful game actions. This training has been found to support both speech (21, 22) and nonspeech (8,

22–24) category learning that generalizes to novel exemplars. The current study tests whether the striatum is recruited during this learning and, if so, whether its engagement is modulated by the presence of coherently structured category input distributions and associated with the hallmark of category learning, the ability to generalize learning to novel instances.

## Results

A core assumption of the videogame is that the utility of sounds in guiding successful gaming action leads to implicit learning about the correspondences among the sound categories and the videogame's visuomotor task demands. In our first-person space-themed videogame, the performance objective is to execute appropriate gaming actions upon the approach of visual aliens. Each alien is associated with a particular sound category defined by four acoustically variable exemplars, and a particular region of the visual space from which it originates. When an alien appears, one sound category exemplar is presented repeatedly until the player executes an action. At higher game levels, the sounds precede the visual appearance of an alien. This creates the opportunity for players to utilize the sound category information to predict the appearance of the specific aliens and execute the appropriate action within the time demands.

We critically manipulated the utility of the sound categories by varying, across two participant groups, the distribution of the exemplars assigned to the aliens (Fig. 1). An experimental group experienced aliens for which the set of copresented sounds was organized into coherently structured auditory categories, with



**Fig. 1.** Overview of the experimental approach. (A) Schematic illustration of the sound categories. Each alien was paired with four sound exemplars. Each exemplar was composed of an invariant low-frequency spectral peak (P1; dashed line) and one of the four possible higher-frequency spectral peaks (P2; either a gray or black solid line). (A, Upper) The four categories learned by the experimental group. (A, Lower) The categories learned by the control group. Each group experienced an identical pair of offset sound categories, defined by a single acoustic dimension (i.e., decreasing vs. increasing in frequency across time). Only the onset-sweep categories differed across groups. For the experimental group, onset categories were organized into two categories potentially linearly separable in a higher-dimensional perceptual space defined by the integration of multiple acoustic dimensions [i.e., steady-state (SS) frequency and P2 onset-frequency locus]. For the control group, onset-category exemplars were randomly drawn from this space (Methods). The scatterplots show the higher-dimensional relationships of the onset-sweep categories for each group, with color indicating the categories. (B, Left) A game screenshot. Each alien creature appears from a consistent quadrant. (B, Right) A schematic depiction of a game trial structure with multiple events. On each trial, one alien appears and a single exemplar from the associated sound category is presented repeatedly. Participants must navigate the videogame to center the alien and take the correct gaming action (23). The trial length (appearance of each alien) depends upon how well participants play the game; thus, it is highly variable within and across participants.

category boundaries characterized by unidimensional or multidimensional spectrotemporal features (the offset- or onset-frequency sweep, respectively). Thus, each of the four aliens was associated with a category of sounds that could, in principle, support generalization through its distributional structure. A control group experienced offset categories identical to the experimental group. However, for the onset categories, exemplar sets assigned to an alien were randomly drawn from the acoustic space that characterized the experimental group onset categories. This created two onset-sweep categories that could not be distinguished from each other on the basis of a linear boundary; however, the onset categories were perceptually distinctive from the offset categories. Thus, item-specific learning of onset-category exemplars is possible but, without distributionally structured input that distinguishes between the onset categories, the ability to generalize onset-category learning to novel exemplars should be restricted to the coarse distinction between onset and offset categories (*SI Appendix, Results and Fig. S1*). This design therefore allows examination of the relationships among behavioral category-learning outcomes, behavior in the videogame task, and involvement of the striatum as a function of the distributional structure of the sound categories.

**Category Distributional Structure Affects Incidental Auditory Category Learning via Videogame Play.** We examined whether participants learned sound categories incidentally from game play, using accuracy assessed in a postscan explicit category labeling task. In this task, listeners categorized both familiar exemplars experienced in the videogame and novel exemplars to test generalization of category learning (Fig. 2).

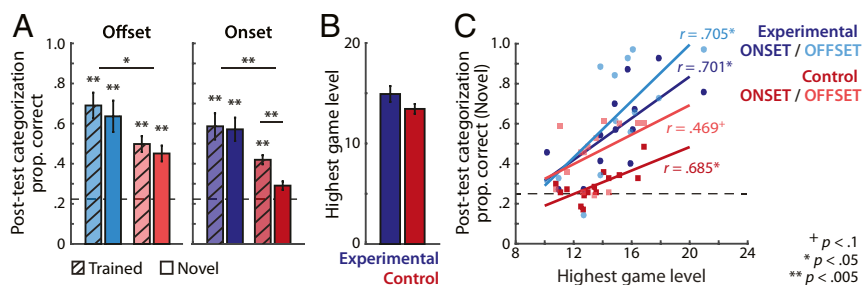
First, we analyzed category learning of the onset categories that critically differentiated the two groups and presented the most difficult category-learning challenge using a mixed-effects logistic model. This analysis revealed significant effects of group, sound exemplar type, and their interaction [Group  $\times$  Exemplar type:  $\chi^2(1) = 9.73, P = 0.0018$ ; *SI Appendix, Table S1, onset*]. Post hoc tests revealed that the experimental group exhibited above-chance generalization of learning to novel exemplars (experimental:  $t_{12} = 5.53, P < 0.0005$ ), categorizing them just as accurately as familiar, trained exemplars ( $t_{12} = 0.30, P = 0.77$ ). In contrast, the control group did not generalize category learning to novel exemplars, leading to significantly poorer categorization accuracy for novel, compared with familiar, exemplars ( $t_{13} = 6.40, P < 0.0005$ ) and overall performance that did not differ from chance (control:  $t_{13} = 1.95, P = 0.073$ ). [The trend-level generalization result for the control group, in contrast, reflected a limited ability to coarsely discriminate between the onset- vs. offset-sweep categories (*SI Appendix, Results and Fig. S1*).]

Next, we analyzed accurate categorization of the offset categories shared across groups. A mixed-effects logistic model analysis of accurate categorization of familiar vs. novel exemplars in the two participant groups revealed main effects of group and sound exemplar type, but no significant Group  $\times$  Exemplar type effect [ $\chi^2(1) = 0.47, P = 0.49$ ; *SI Appendix, Table S1, offset*]. Overall, the experimental group exhibited more accurate categorization than the control group ( $M_{\text{diff}} \pm \text{SE} = 0.19 \pm 0.078$ ), and familiar exemplars were more accurately categorized than novel exemplars ( $M_{\text{diff}} \pm \text{SE} = 0.05 \pm 0.019$ ). Generalization to novel exemplars was similar across groups, and above-chance for each group (Fig. 2A; experimental:  $t_{12} = 4.94, P < 0.0005$ ; control:  $t_{13} = 5.14, P < 0.0005$ ).

Since generalization of learning to novel exemplars is a hallmark of category learning, these results indicate that both groups incidentally learned the offset categories. In contrast, robust learning of the onset categories was shown only in the experimental group that experienced coherent distributional structure [see *SI Appendix, Table S1* for a full model for a significant Group  $\times$  Sound Category type  $\times$  Exemplar type effect;  $\chi^2(1) = 6.61, P = 0.010$ ]. Further investigations of categorization response error patterns revealed that the experimental group categorized the four sound categories, whereas the control group coarsely discriminated the onset- vs. offset-category types (*SI Appendix, Fig. S1*).

In previous research using the same paradigm, the highest videogame level reached during training (an indirect measure of learning) has been positively correlated with explicit posttest categorization accuracy (23–25). In the present conditions of videogame training within the scanner, there was no significant group difference in game performance ( $t_{25} = 1.63, P = 0.12$ ; Fig. 2B) or in the duration of self-paced game training trials (*SI Appendix, Fig. S2*). Note that although auditory category learning supports game success, it is not essential; it is thus noteworthy that there was a reliable association of videogame performance with accuracy in generalizing category learning to both novel onset exemplars (Fig. 2C; experimental:  $r = 0.701, P = 0.007$ ; control:  $r = 0.685, P = 0.007$ ) and novel offset exemplars (experimental:  $r = 0.705, P = 0.007$ ; control:  $r = 0.469, P = 0.098$ ).

**Involvement of the Striatum During Incidental Auditory Category Learning Depends upon Category Distributional Structure.** Previous studies have emphasized the involvement of the striatum in perceptual category learning (e.g., refs. 4, 12, 26, and 27) and nondeclarative, implicit learning (e.g., refs. 28–31) using explicit classification tasks. Thus, we investigated whether the striatum is recruited for incidental learning of sound categories, and whether the recruitment differs and varies as a function of the exemplar distribution (experimental, control) that drives differences in behavioral learning outcomes.



**Fig. 2.** Behavioral performance. (A) Proportion correct categorization of offset- and onset-category exemplars in the posttest. Hashed bars indicate mean performance across the trained exemplars experienced in the videogame. Solid bars indicate mean performance for novel exemplars withheld from training (i.e., generalization) in the experimental (blue) and control (red) groups. (B) Group average in-scanner videogame performance, as measured by the highest game level achieved. (C) Correlation of the videogame performance (highest game level achieved) and categorization accuracy for the novel offset- and onset-category exemplars. Pearson's correlation  $r$  values are shown. Dashed lines in A and C indicate chance level (0.25) performance. Error bars indicate  $\pm 1$  SEM. See *SI Appendix, Fig. S1* for further analysis of behavioral error pattern differences across the experimental and control groups.

Given the unique nature of the videogame task, we estimated the neural responses during game play as a time-dependent change of blood oxygen level-dependent (BOLD) magnitude from the onset of a game trial while controlling for variations in trial duration (Fig. 1*B* and *Methods*). We first identified regions within the striatum that exhibited different BOLD activation time courses between the two participant groups in response to the audiovisual events in the videogame (Fig. 1*B*). Within a striatum anatomical mask (*Methods*), we found that the left caudate body and bilateral putamen (Fig. 3*A* and *SI Appendix, Table S1*) consistently exhibited a significant Group  $\times$  Time Course effect across three runs, even though the two groups did not differ in overall game performance (Fig. 2*B*) or in game trial duration (and, thus, timing of feedback in the game; *SI Appendix, Fig. S2*). On average, the experimental group recruited these striatal regions to a greater degree than the control group (Fig. 3*B*). It is of note that this effect is not confined to the striatum, the focus of our predictions; in fact, it involves other cortical regions as well (*SI Appendix, Table S2*). Interestingly, no brain regions were sensitive to the different sound category types (onset vs. offset); rather, the BOLD responses to the two category types were highly similar in both groups during game play (*SI Appendix, Fig. S3*). Thus, we used the overall BOLD response in subsequent analyses, collapsing across the two category types.

Next, we sought to understand the relationship of striatal activity to behavioral indicators of auditory category learning. Using multiple linear regression, we tested whether the magnitude of BOLD activity in the striatal clusters (*SI Appendix, Table S1*) differentially predicted participants' behavioral performance across the two groups. Turning first to the onset categories that differentiated the groups, we observed no main effects ( $P_s > 0.19$ ) but found a significant Group  $\times$  Striatal Activation effect ( $\beta = 0.70, t = 2.37, P = 0.027$ ; *SI Appendix, Table S4*), indicating that the two groups exhibited a different relationship between striatal activity and the behavioral generalization of onset-category learning in the explicit posttest (Fig. 4*A, Left*). Post hoc correlation analyses performed on each group revealed a trend toward a positive relationship in the experimental group ( $r = 0.54, P = 0.058$ ) but no relationship in the control group ( $r = -0.37, P = 0.20$ ). In contrast, there were no significant main or Group  $\times$  Striatal Activation effects for the offset categories shared across the groups (*SI Appendix, Table S4*), which indicates that the relationship between the striatal activity and the behavioral generalization of offset-category learning did not differ across the groups.

A similar pattern was apparent between striatal activation and in-game performance (Fig. 4*A, Right*). The Group  $\times$  Striatal Activation effect was significant ( $\beta = 0.91, t = 2.58, P = 0.017$ ), but there were no main effects of Group ( $\beta = -0.32, t = 1.35, P = 0.19$ ) or

Striatal Activation ( $\beta = -0.069, t = 0.25, P = 0.81$ ) in predicting videogame performance. Post hoc analyses revealed that the magnitude of striatal activation was strongly positively correlated with videogame performance only in the experimental group ( $r = 0.74, P = 0.004$ ; control:  $r = -0.072, P = 0.81$ ).

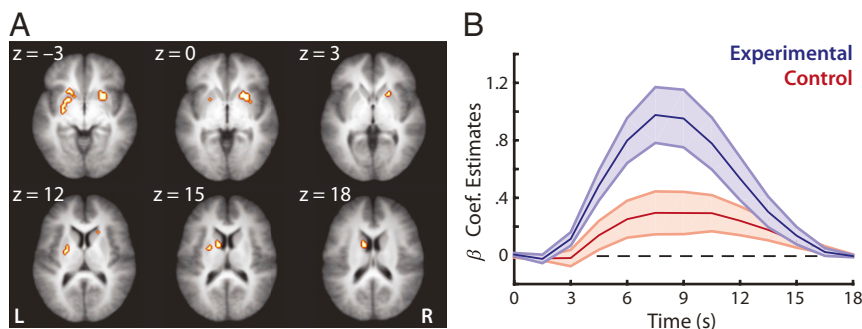
In all, a set of posterior striatal regions exhibited a significant Group  $\times$  Time Course effect in response to the audiovisual events experienced in the videogame (Fig. 3), and the magnitude of striatal activation was a better predictor of auditory category learning in the experimental group than the control group (Fig. 4*B*).

### A Striatum–Left Superior Temporal Sulcus Auditory Loop Contributes to Learning.

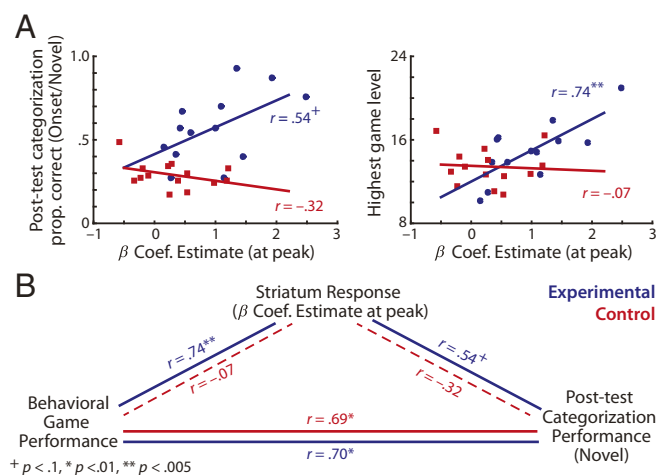
A prior study found learning-related activation increases in a putatively speech-selective region of the left superior temporal sulcus (l-STS) after videogame training of distributionally structured (but not unstructured) nonspeech (8). However, the pre- vs. posttraining fMRI design in this prior study precluded examination of the learning mechanisms and associated brain systems that drive these cortical changes. In this context, we examined whether the striatum was functionally connected to the l-STS during videogame play, and whether this connectivity was related to behavioral measures of auditory category learning in our experimental and control groups who experienced onset categories with different degrees of distributional structure.

Using localizer data, we first identified a speech-selective region of interest (ROI) for each participant within the l-STS. First, as in Leech et al. (8), we found that the experimental group exhibited robust activation during game training in similar regions in the l-STS (Fig. 5, *Left*), and this l-STS activation was a better predictor of onset-category generalization performance in the posttest in the experimental, compared with the control, group (*SI Appendix, Fig. S4*). With these localized l-STS ROIs, we next performed connectivity analysis; our striatal regions (Fig. 3*A* and *SI Appendix, Table S2*) served as seeds, and functional correlations were computed and extracted from individually defined speech-selective ROIs in the l-STS (Fig. 5). In each group, the striatal regions were functionally connected with the speech-selective ROIs during game training (one-sample  $t$  tests against 0;  $t_s > 9.37, P_s < 5.14 \times 10^{-7}$ ), and each group exhibited a similar degree of striatal connectivity with the l-STS ROIs, as well as with other areas across the whole brain (i.e., no brain area exhibited significant group differences in connectivity).

This established a context in which we could examine whether the degree of functional connectivity between striatal and speech-selective l-STS tissue is differently related to behavioral performance in the two groups. To this end, we performed a multiple linear regression analysis examining the generalization of the onset



**Fig. 3.** Activation of the striatum across the experimental and control groups during videogame training. (*A*) Regions within the striatum exhibiting a significant Group  $\times$  Time Course effect in response to game audiovisual stimulus events. The highlighted regions are a union of active striatal clusters across the three functional runs (caudate body and putamen; *SI Appendix, Table S1*). (*B*) The average activation time course of the striatal regions active in videogame play. Shaded regions indicate  $\pm 1$  SEM. Post hoc tests revealed greater BOLD activation in the experimental than control group ( $t_{25} > 2.23, P_s < 0.035$  from 4.5 to 10.5 s from the start of each game trial).



**Fig. 4.** Relationship between behavioral measures and striatal activation. (A) Correlation between the activation of striatal clusters during game play and generalization of category learning to novel onset-category exemplars in an explicit posttraining categorization test (Left), and in-scanner videogame performance (Right). Data points and regression lines of the experimental and control groups are indicated by blue and red, respectively.  $\beta$ -Coefficient estimates indicate the peak activation of the striatum during game play shown in Fig. 3B. (B) Summary of the relationships among behavioral game performance, posttraining categorization, and striatal activation during game play. Pearson's correlation  $r$  values are derived from an individual, post hoc correlation analysis following the significant Group  $\times$  Striatal Activation interaction effect from a multiple linear regression analysis predicting corresponding behavioral performance measure.

categories in the posttest based on each group's functional connectivity between the striatal and I-STs ROIs; this analysis revealed a significant main effect of Group ( $\beta = 1.44, t = 3.76, P = 0.001$ ) but no effect of Striatal Connectivity ( $\beta = 0.038, t = 0.22, P = 0.82$ ). Most importantly, we also found a significant Group  $\times$  Striatal Connectivity effect ( $\beta = -0.84, t = 2.30, P = 0.031$ ). This key finding suggests that the two groups exhibited a different relationship between the striatal connectivity to I-STs and behavioral generalization (Fig. 5). Post hoc correlation analyses revealed that the functional connectivity between the striatum and I-STs in the experimental group was marginally related to onset-category generalization ( $r = -0.52, P = 0.071$ ), but there was no relationship in the control group ( $r = 0.11, P = 0.700$ ). On the contrary, the same analysis of offset-category generalization revealed neither significant main effects ( $t_s < 1.18, P_s > 0.25$ ) nor a Group  $\times$  Striatal Connectivity interaction ( $\beta = 0.097, t = 0.19, P = 0.849$ ). The predictive relationship between the striatal connectivity with I-STs and offset-category generalization did not differ across the groups.

Our main focus was to investigate the connectivity between the striatum and speech-selective I-STs but, given the nature of the videogame task, we also performed a secondary analysis to explore the connectivity between the striatum and the localized visual and motor regions (SI Appendix, Methods). For visual regions, we found patterns of connectivity to behavioral measures of category learning similar to those of I-STs. For bilateral motor regions, the connectivity-behavioral relationship was weaker (SI Appendix, Fig. S5).

## Discussion

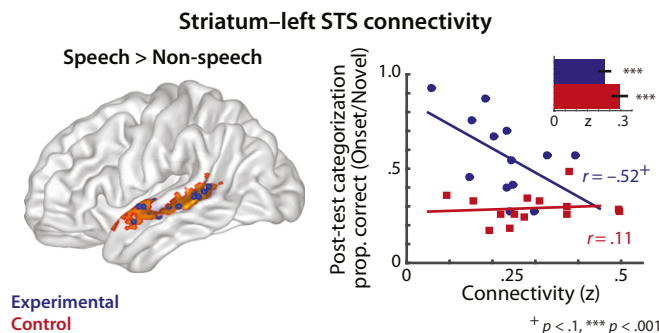
What we know about the neural mechanisms that support category learning has come largely from studies of visual category learning under explicit training. However, category learning outside the laboratory often involves learning under conditions in which the relationship of category exemplars to behaviorally relevant behaviors and events is learned incidentally without overt feedback or explicit category decisions. Further, we know that auditory category learning, such as the learning of speech categories, is an important

human skill. By using a novel, videogame-based approach, we investigated whether the striatum is engaged in incidental auditory category learning, and whether it is sensitive to the distributional structure of sound categories even when attention is not explicitly directed to the sounds. In fMRI, participants played a videogame in which sound category exemplars were predictive of successful actions. For an experimental group, each alien was associated with a category of sounds defined by exemplars with distributional regularity in perceptual space. In contrast, for a control group, the exemplars for two aliens lacked an organized distributional regularity across exemplars, thereby creating an auditory environment with less statistically coherent structure associated with task-relevant behaviors and events.

Our findings can be summarized as follows: (i) Listeners incidentally learned auditory categories in the course of videogame play, even without explicit categorization of, or directed attention to, the sounds, and generalized that learning to novel exemplars when the categories possessed distributional regularity; (ii) the striatum was engaged in a learning-dependent manner during this incidental learning; (iii) functional connectivity between the striatum and left STS tissue involved in complex sound categorization (8) was modulated as a function of incidental category learning; and (iv) the modulations of striatal activity and connectivity to the left STS differed across the experimental and control groups, indicating the sensitivity of the striatal learning system to coherently structured stimulus input. At the broadest level, our results demonstrate that the striatum contributes to incidental sound category learning even when behavior is directed at navigating a rich, multimodal environment and not focused on sound categorization per se.

## The Striatum Facilitates Learning Distributions of Category Exemplars.

Recruitment of the striatum may be essential in accomplishing effective sound category learning across distributions of sounds that are difficult to acquire through unsupervised learning (15, 16). For



**Fig. 5.** Functional connectivity between the posterior striatum and speech-selective left superior temporal sulcus tissue. (Left) I-STs tissue was identified as exhibiting greater BOLD activation for hearing speech compared with nonspeech sounds (i.e., speech: English words and syllables vs. nonspeech: semantically matched environmental sounds and sound exemplars from the videogame) in a separate localizer task before videogame training. The group-based speech > nonspeech contrast mask is shown in orange, and individually defined speech-selective ROIs are shown as blue spheres. (Right) Correlations between generalization of onset-category learning in the posttest and the extent of functional connectivity between the striatum to the localized I-STs ROIs on a single-subject basis. Behavioral categorization performance (y axis) represents generalization performance for categorizing novel onset-category exemplars. The bar graph (Inset) shows average connectivity for each group from the striatal seeds to individually defined speech-selective ROIs. The functional connectivity measure is expressed in Fisher's  $z$ -correlation coefficient. The correlation  $r$  values are derived from post hoc correlation analyses, following upon a significant Group  $\times$  Striatal Connectivity interaction effect from a multiple linear regression analysis. Error bars indicate  $\pm 1$  SEM.

instance, passive listening to difficult nonnative speech categories does not recruit the striatum, and produces few or no learning gains (12, 13, 32). In contrast, overt categorization tasks with explicit feedback elicit robust activity in the striatum and yield substantial learning (12, 33). Similarly, although the onset categories of the present study were learned in the context of short-term incidental videogame training, they are not learned via passive listening (23, 34).

The difference may be that the active incidental videogame training introduces opportunities for the distributional regularities of sound category exemplars to align with behaviorally relevant actions and events. Prior research involving overt responses and associated feedback has demonstrated that striatal recruitment is particularly robust in tasks that involve goal-directed actions, a sense of agency over contingent outcomes, and information that provides performance feedback (35, 36). The present videogame possesses these task features; even without overt category decisions, the active nature of the videogame task encourages sound-based predictions to drive behavior, with associated positive (e.g., points earned) and negative (e.g., alien death) outcomes. This can explain the similar degrees of striatal connectivity found in the two participant groups across auditory, visual, and motor regions (Fig. 5 and *SI Appendix, Fig. S5*).

Critically, the consistent relationship of category exemplars to a specific alien (and action outcomes associated with the alien) relates the multiple, acoustically variable exemplars from a sound category to consistent behaviorally relevant aliens and actions. These characteristics may more successfully drive category learning than unsupervised learning (i.e., passive exposure). We speculate that incidental learning can outperform unsupervised learning by drawing upon reinforcement learning mechanisms involving the striatum. We base this on evidence that the posterior striatum accumulates information about the reward history of stimuli to yield value-based representations across the perceptual space that can guide optimal responses (37) and which may, via corticostriatal loops (5), hasten the alteration of cortical representations in the left STS (*SI Appendix, Fig. S6*).

**The Striatum Is Sensitive to Distributional Structure in the Stimulus Input.** Prior research has emphasized the role of the striatum in gradual learning of stimulus–response relationships to rewarding outcomes, often in the context of simple stimuli that vary in reward value (31, 37, 38). The present results highlight that mappings of coherently structured category exemplars also accumulate value through striatal learning. To understand this point, it is helpful to recall that the stimulus–response mapping was equivalent across the experimental and control groups; all participants experienced a perfectly consistent mapping between sound exemplars and a specific game event (alien). What differed was the relationship of these exemplars to one another in perceptual space; onset-category exemplars possessed distributional regularity for the experimental, but not the control, group. Although the groups had similar game performance and engagement (Fig. 2*B* and *SI Appendix, Fig. S2*), and all exemplars had a consistent alien–action mapping across learning, the distributional structure of the onset categories significantly affected the course of learning. Categorization performance of the control group was significantly poorer than the experimental group even for the offset categories, which were identical across groups and can be discriminated without training due to the unidimensional cue that differentiates them (23, 34). Likewise, the control group engaged the striatum to a lesser extent than the experimental group, for both the onset and offset categories. These patterns potentially highlight how the distributional regularities in the category stimulus input (even in a subset of the input) can differently leverage the striatal system that accumulates behaviorally relevant valuation of the stimulus input during learning.

The importance of a coherent distributional structure for learning outcomes—especially generalization—may be related to the way in which reinforcement learning proceeds via gradual accumulation of stimulus–response relationships according to their reward value (*SI Appendix, Fig. S6*). The striatum has great potential for information compression, which might be crucial in acquiring representations that support generalization to new exemplars (1). Given the high convergence ratio from cortical input to the striatum (39) and to basal ganglia output (40–42), corticostriatal loops can carry out neural computations required in acting as a pattern detector (39, 43). In this way, the striatum is well-suited to learning distributionally structured exemplars and may be less efficient when a coherent distributional structure is absent, or less robust as in the control group.

We propose that exemplars with distributional regularity, as in the experimental group, will tend to accumulate reward value in a common region of perceptual space, thereby building a coherent “value map” across encounters that can be used to more effectively drive future reward–yielding actions and, ultimately, allow for the generalization of this learning to novel exemplars that fall in a similar region of perceptual space (*SI Appendix, Fig. S6*). Even when the distributional regularity of category exemplars is not perceptually apparent, as is the case for the onset-category sounds (23, 34), striatal engagement can capitalize upon the distributional structure, in turn yielding better learning of objects and events that align with positive behavioral outcomes.

In the same manner, for the control group, the onset-category exemplars were scattered in perceptual space (without robust distributional regularity), providing less opportunity to relate a consistent region of perceptual space to a specific alien, and to the action needed to produce a rewarding outcome. This is consistent with poorer behavioral learning and reduced striatal engagement for the control group, observed especially for the unstructured onset categories. Although overall recruitment of the striatum was reduced for the control group compared with the experimental group, the groups exhibited proportionally similar degrees of generalization for the structured, offset categories (relative to categorization of familiar exemplars). Moreover, where there was distributional regularity for the control group (offset categories), the relationship of striatal activation and striatal connectivity to the speech-selective I-STS area, and to behavior, was similar to that observed for the experimental group. Therefore, we conclude that the striatum participates in learning categories defined by exemplars with distributional regularity that are aligned with value-marking behavioral outcomes.

**The Posterior Striatum Is Recruited During Successful Incidental Learning.** Research has emphasized that different subregions within the striatum have distinct functional roles (1, 17). The head of the caudate nucleus interacts with the prefrontal cortex, and is known to be involved in executive processes (4, 44–46); thus, the caudate head has been implicated in category learning that can be accomplished by selective attention to specific input dimensions (2, 3). In contrast, the posterior striatum—comprising the caudate body, caudate tail, and putamen—has been widely implicated in nondeclarative, implicit learning (19, 20, 47). Similarly, in the context of perceptual category learning, numerous studies have demonstrated the engagement of the posterior striatum when category learning requires integration of information across input dimensions at a predecisional stage (2, 26, 48, 49). This is thought to be accomplished via establishment of a direct motor–response mapping that bypasses frontally mediated explicit or executive processing (6, 29, 50).

This traditional conceptualization would predict that the onset categories (requiring integration across input dimensions) and offset categories (unidimensional up- or downsweep) would engage distinct striatal subsystems for category learning. However, in contrast, we observed posterior striatal engagement irrespective of

category type (*SI Appendix, Fig. S3*). Further, the traditional conceptualization would predict that the onset categories (requiring integration across input dimensions) would first engage anterior striatal learning and then gradually shift to posterior striatal learning (16, 48). Here, we demonstrate the engagement of the posterior striatum from the earliest phases of training (*SI Appendix, Table S1*). One possibility is that when sound category exemplars are a part of a rich, multimodal incidental learning environment, the task complexity may draw attention away from the sounds, diminishing selective attention strategies better aligned with the anterior striatum. The incidental nature of the videogame task also means that participants were not informed about the existence of categories and were not making overt category decisions; this is in contrast to explicit training tasks that have served as the empirical basis for theoretical models that posit distinct, and competitive, roles for anterior and posterior striatal systems. Moreover, since the videogame involved only one exemplar per trial under incidental conditions, there was a strong pressure to gradually accumulate evidence, including reward value, incidentally across alien encounters. This aligns with functional roles associated with the posterior striatum (e.g., refs. 37 and 51) and, as highlighted above, might be comparatively less optimal for the control group, since the sound exemplars possess less distributional structure compared with the experimental group.

Overall, the present results indicate that the posterior striatum is not limited to the learning of complex categories that require integration across input dimensions. Here, learning of the offset categories that are easily defined by a single, verbalizable input dimension to which selective attention can be directed also engaged the posterior striatum, at least when the sounds were embedded in an incidental task with coherent exemplar regularity that aligned well with the behavioral goal. Interestingly, our previous behavioral work on adult listeners suggests that the learning of nonnative speech categories within the videogame is more efficient (21) than extensive training using explicit tasks (e.g., refs. 52 and 53). Potentially, the learning advantages of the incidental videogame task over explicit training might derive from the early engagement of learning mediated by the posterior striatum.

Extending our results to the consideration of naturalistic speech learning, the learning capacity of the posterior striatum might help listeners become attuned to functionally relevant linguistic units (e.g., words and phonemes) experienced in complex, multimodal contexts that involve contingent rewards. For example, this could include naturalistic social interactions that involve speech and reinforcing reactions (such as a smile, eye gaze, or vocal response) (54–56). The gradual accumulation of speech interactions with associated rewards could guide learning of consistent (and predictive) correspondence in the multimodal input (e.g., multiple instances of speech input and the visual referents, such as hearing “bear” while seeing a bear doll vs. “pear” while seeing a pear fruit). As such, the incidental nature of the videogame task might closely model the complexity of naturalistic and interactive learning environments.

**Contribution of the Auditory Corticostriatal Loop During Category Learning.** Our results indicate that the posterior striatum is sensitive to the coherent distributional structure of auditory categories that listeners incidentally learn during videogame play. Previous studies have shown that the posterior striatum indeed interacts with sensory cortical regions to support perceptual category learning (5, 57). Thus, we also expected the posterior striatum to exhibit evidence of interaction with cortical regions implicated in auditory category perception and learning. We were particularly interested in the I-STS, because activation in this area has been linked to general expertise with complex sound and phonemic categories (7, 58). Even more directly, the I-STS activity observed in the current study is similar in location to the learning-dependent activity increases observed following videogame training with the same set of

sound categories as those learned by the experimental group (8). Here, we observed that during learning across videogame play, the experimental (but not the control) group recruited the I-STS ROIs with activity levels that predicted behavioral categorization performance in the posttest.

Thus, we predicted that the posterior striatum is engaged in auditory category learning through a functional circuit involving the I-STS. Consistent with this prediction, seed-based connectivity results revealed that functional connectivity between striatal and I-STS ROIs predicted a behavioral measure of category learning only for the experimental group (Fig. 5). This suggests that learning-related modulation of connectivity between the posterior striatum and the I-STS may signal the engagement of a corticostriatal loop that can drive learning of statistically coherent category input structure.

Of note, the magnitude of striatal connectivity to the I-STS (and other visual and motor regions relevant to multimodal videogame training) did not differ between groups. This might suggest that the incidental videogame task engaged the striatal learning system for each group, widely broadcasting and receiving from various cortical regions relevant to successful videogame performance. However, note that we observed a learning-related decrease in striatal connectivity only for categories with distributional exemplar structure (Fig. 5 and *SI Appendix, Fig. S5*). Whereas the striatum may broadcast learning signals to cortical regions broadly relevant to the incidental task, the utility of this signal for category learning was contingent upon the alignment of videogame actions and outcomes with categories possessing distributional regularities.

## Conclusion

This study investigated the contributions of the striatum to incidental auditory category learning. We examined this issue by utilizing a videogame task (23) that presented an opportunity to learn auditory categories in a rich, multimodal environment without explicit feedback about overt category decisions. The results show that the posterior striatum—especially the caudate body and putamen—contributes to the incidental learning of functionally relevant sound categories. Moreover, the posterior striatum interacts with a specific cortical area, the left superior temporal sulcus, which has previously been implicated in complex auditory category perception and learning (8). By creating different learning contexts that manipulated the functional utility of sound category information, we demonstrate that the nature of the distributional structure of sound categories greatly impacts how the striatum is recruited and interacts with the auditory association cortex. Together, the present results provide evidence for a functional role of the posterior striatum in incidental category learning. This finding may particularly be relevant in understanding speech category learning, which proceeds incidentally without explicit feedback (15, 21). In the context of speech category learning, modulation from attentional and motivational factors (55) and contingent extrinsic reinforcers like social cues (54) may play an important role in drawing striatal systems online to support and facilitate incidental category learning.

## Methods

**Participants.** Thirty right-handed native-English speakers were assigned to one of the two groups (Fig. 1). Two participants were removed from data analysis due to excessive head movement during scanning, and one was removed due to an incidental neurological finding. This resulted in 13 participants (6 females) in the experimental group, and 14 (6 females) in the control group (mean age 22.7 y). Participants gave informed consent and were paid \$60 upon completion. All reported normal hearing and no history of neurological impairments or communication disorders. The study procedure was approved by the University of Pittsburgh.

**Auditory Stimuli and Group Assignment.** The stimuli were identical to those used in previous studies (8, 23–25). The present study involved a total of six auditory categories, but each participant experienced four categories

among the six. All sounds were artificial, nonspeech sounds composed of a steady-state frequency and a frequency transition in each of two spectral peaks (P1 and P2; Fig. 1). The exemplars within and across categories were acoustically similar.

Following Wade and Holt (23), the two groups of participants (experimental and control) were defined by the differences in the to-be-learned sound categories. In brief, the groups shared two offset-sweep categories. For these categories, the P2 frequency either increased or decreased across time beyond the onset frequency. This provided a single acoustic dimension that unambiguously signaled category membership (Fig. 1).

For the remaining onset-sweep categories, P2 frequency increased or decreased over the initial 150 ms of each sound and then remained at a steady-state frequency for 125 ms. The two categories experienced by the experimental group had different P2 onset-sweep loci that varied across exemplars. Thus, no single acoustic characteristic was sufficient to determine category membership, but there was higher-order structure across multiple acoustic dimensions (i.e., sound steady-state frequency and P2 onset locus) that defined the categories (Fig. 1A, Upper). [These categories have different P2 onset loci (high vs. low in frequency), but the onset locus does not serve as an a priori cue to distinguish the two categories. Prior studies have shown that while the two offset-sweep categories can be discriminated without any training, the onset categories are indistinguishable by naive listeners (23, 34).]

In contrast, the two onset-sweep categories experienced by the control group did not have this higher-order structure (Fig. 1A, Lower). These category exemplars were drawn from the same acoustic space as the experimental group onset-sweep categories; however, there was no constant P2 locus associated with a category, and so no higher-order structure. Therefore, the two onset-sweep categories experienced by the experimental vs. control group differed in whether they could be described by the presence or absence of complex statistical structure—namely, whether there exists a separable category boundary between them on the basis of spectrotemporal information integrated across dimensions (23).

These stimuli were designed to mimic spectrotemporal characteristics of speech. The two spectral peak components were analogous to formant frequencies in speech. The frequency transition patterns of the offset- and onset-sweep categories attempted to model vowel-consonant and consonant-vowel syllables; the steady-state phase modeled the vowel component, whereas the rapid transition modeled the consonant component. Each category consisted of 11 sound exemplars: Four were experienced in game training (Fig. 1A), and seven were withheld to test generalization in a behavioral posttest. All exemplars were matched in root-mean-square amplitude. Each of the four categories was paired with one of the four alien characters.

**Experimental Procedure.** Participants completed a single 1.5-h fMRI session. The fMRI session included a localizer task, videogame training across three functional runs, and an auditory habituation task (which is not reported here due to excessive head movements). In fMRI, sound stimuli were presented through MR-compatible headphones (Silent Scan Audio System; Avotec), and right- and left-hand motor responses were recorded using an MR-compatible number pad (Mag Design and Engineering) and response glove (Fiber Optic Button Response System; Psychology Software Tools), respectively. After the videogame training in fMRI, participants completed an explicit sound category labeling task in a quiet behavioral testing room. All participants experienced the same experimental protocol and the same offset sound categories, except for the distributional structures of onset-sweep categories.

**Localizer Task.** All participants first completed a localizer run with an aim of localizing activity within the I-STS area, identified by Leech et al. (8) as selective to learning-related changes after incidental sound category learning in the videogame. There was a total of 96 auditory trials, on which participants heard either natural speech (English word or syllable) or nonspeech sounds (environmental sounds or nonspeech sounds from the videogame). The English words and semantically matched environmental sounds (e.g., spoken word “pouring” vs. the sound of pouring a drink) were selected from the localizer used by Leech et al. (8). To ensure that participants remained alert, the task also had 72 active, visuomotor trials, in which listeners pressed button keys associated with the color of the visual stimuli. Each trial lasted 1.5 s (see *SI Appendix, Methods* for further details).

**Videogame Training.** After the localizer run, participants played the videogame across three runs (12 min per run). During this incidental training, participants actively navigated a virtual environment, and encountered four visually distinct aliens. Each alien was associated with a particular category of

sounds and a game action (shooting or capturing). Each alien appeared consistently from a distinct quadrant, with the exact position randomly jittered (Fig. 1B). Each time an alien appeared and progressed forward on the screen, one randomly selected sound exemplar from the associated category was repeatedly presented until the participant acted on the alien and it disappeared. Players freely navigated the game environment to orient toward the alien using the right hand and took appropriate game actions (shoot/capture) with the left hand. Depending upon the success or failure in taking appropriate game actions on the alien, a 500-ms visual feedback was presented (Fig. 1B, Right). Note that this feedback was based on success in performing correct game actions and not linked to sound category decisions per se.

As the game advanced to higher levels, aliens moved forward at a faster speed, and originated farther out from the center of the screen, eventually reaching a point where players could hear the alien before seeing it. Since the sound category membership perfectly predicts the alien and its quadrant of origin, the sound category exemplars have functional value in signaling appropriate game actions. Thus, success in the game is incidentally linked to the ability to generalize across acoustic variability in category exemplars to support the rapid execution of appropriate action. We evaluate this with the metric of highest game level achieved in game play (23).

Participants played the game in a self-directed manner with game difficulty and speed determined by their own behavior. As a result, there was no fixed trial structure, as is typical of imaging experiments. For the purposes of analysis, we defined the onset of “game trials” to be the start of a repeatedly presented auditory category exemplar (which were temporally synced with the onset of functional image acquisition). A randomly selected intertrial interval (0 to  $-4.5$  s) and an additional delay ( $<1.5$  s) separated game trials (see *SI Appendix, Methods* for more details).

Participants were instructed only on how to play the game. They were not informed about the sound categories or their significance in the game, and synthetic game music was played in the background (23). Thus, unlike traditional category-learning tasks, the game training did not involve overt attempts to discover the dimensions diagnostic to category membership or explicit categorization decisions. Further, the task provided an indirect form of feedback, as correct prediction of an alien based upon its emitted sound should enhance success in the game. In all, the videogame task presents an inherently complex environment in which participants experience rich correlations among multimodal cues that support success in the game and incidental learning of sound categories.

Afterward, all participants went through two additional functional runs of an auditory habituation task (not reported due to excessive head motion), which lasted about 20 min.

**Behavioral Posttest.** After the scanning session, participants completed a posttraining overt category labeling task to assess category learning and generalization in a quiet behavioral testing room. To refresh the mapping of the sounds and the alien identity, participants first played the videogame on a laptop for 15 min before the posttest. On each trial of the posttest, participants saw the four alien characters in fixed positions on the screen and heard a randomly selected sound category exemplar, which was repeatedly presented. Participants indicated which alien was associated with the sound exemplar within a 5-s window. In addition to the four familiar category exemplars experienced in the videogame, participants also responded to seven novel exemplars from each category that were withheld from the videogame training to test generalization of category learning. Each sound was presented five times (220 trials). There was no feedback.

**Behavioral Data Analysis.** Two behavioral learning measures were used. One measure was participants’ game performance during training in the scanner. This measure was quantified as the mean of the highest levels attained across the multiple games played by each participant (23). The other measure was participants’ explicit sound categorization performance assessed in the posttest following the in-scanner videogame training. Our main interest was to examine the extent of generalization of category learning to novel exemplars assessed in the posttest (i.e., after the category learning that took place in the videogame). See *SI Appendix, Methods* for details of the statistical tests.

**Functional Image Acquisition and Preprocessing.** Brain images were acquired on a Siemens Allegra 3.0 Tesla scanner at the University of Pittsburgh. Preprocessing and analysis were done using Analysis of Functional Neuroimages (AFNI) (59). Preprocessing steps included slice-time correction, spatial realignment to the first volume, and normalization to Talairach

space (60). All imaging parameters and preprocessing details can be found in *SI Appendix, Methods*.

**Localization of Speech-Selective ROIs.** As our previous study exhibited pre vs. post videogame training effects in the speech-selective cortical region (8), we used the data from the localizer task to identify speech-selective ROIs. The localizer task general linear model (GLM) included a separate regressor for each trial event (auditory: speech vs. nonspeech; visual: alien vs. Gabor patch; motor: left vs. right motor responses), which were modeled at the onset of each trial, convolved with a canonical gamma-variate function (GAM; AFNI).

As individuals greatly vary in their brain anatomy and functional recruitment, we localized candidate speech-selective regions on a single-subject basis. Within the group-defined mask (at voxel-wise  $P < 0.05$ ; cluster-corrected threshold of 263 voxels), we identified each individual participant's peak voxel exhibiting speech > nonspeech contrast. Around each individual's peak activation voxel, we created a speech-selective ROI for each participant using a 2.5-voxel-radius sphere (~85 voxels). Our primary interest was in the auditory ROI, but for our exploratory analysis, we also identified the ROIs reflecting the visual and motor components using a similar approach (*SI Appendix, Methods and Fig. S5*).

**Videogame Task GLM.** Unlike traditional imaging tasks, the videogame task does not enforce any consistent trial structure or duration. Trial duration was entirely driven by players' performance (see *SI Appendix, Fig. S2* for the distributions of game trial duration). Without any restrictions on when or how to make motor responses, participants freely aimed and executed game actions on aliens using both hands while continuously seeing and hearing aliens. Because of this unusual task design, we define "trial" to be the duration from the onset of a new alien's sound presentation to the receipt of feedback after completing a game action (Fig. 1*B*). Within each trial, there were three consistent game-related events: videogame audiovisual stimulus presentation, motor responses, and feedback, which were separately modeled in the GLM. Our main interest was in the BOLD response to videogame audiovisual stimulus presentation events. Given the continuous and immersive nature of the videogame task, we estimated BOLD responses during the game training using the finite impulse response (FIR) function as a change in BOLD magnitude across time from the game trial onset (i.e., sound onset) up to 18 s. The duration of the stimulus presentation event varied according to the variable self-paced trial length in the game, so a parametric amplitude-modulated regressor that modeled trial duration was entered as a trial-wise covariate to account for variable game trial lengths (see *SI Appendix, Methods* for more details of the GLM).

**Response of the Striatum During Videogame Training.** The central objective of this study was to investigate the role of the striatum in auditory category learning and its sensitivity to the presence of a coherent distributional structure of the input. This analysis was designed to address three main questions: (i) Are regions within the striatum involved during the videogame training; and, if so, do the BOLD responses differ between the two groups of participants? (ii) Are the response magnitudes from the striatum regions correlated with behavioral measures of auditory learning; and, if so, do the patterns of correlation differ between the two groups? (iii) Are regions within the striatum functionally connected to the speech-selective auditory ROIs; and, if so, does the functional connectivity exhibit a different relationship to behavioral measures of auditory category learning between the two groups?

To this end, we created an a priori striatum anatomical mask by using the high-resolution anatomical atlas within AFNI (the TT-atlas). The anatomical atlas provides a parcellation of the striatum into 10 subregions: the left and right head, body, and tail of the caudate nucleus, the left and right putamen, and the left and right nucleus accumbens. Our striatal mask, which was 704 voxels in size, was defined as the union of these striatal subregions. Within this striatal mask, a voxel-wise mixed-design ANOVA was conducted on the videogame audiovisual event data of each functional run; using the weights from the FIR modeling gaming events, the  $\beta$ -estimate of change across time (i.e., time course from 0 to 18 s with respect to the onset of a new game trial) served as a within-subjects factor, group (experimental vs. control) served as a between-subjects factor, and subject served as a random factor (e.g., refs. 12, 35, 45, and 61). Activation clusters within the striatum were defined at a voxel-wise threshold of  $P < 0.005$  and corrected for multiple comparisons at  $\alpha = 0.05$  (3dClustSim by AFNI ver. 17.0.00). We also performed the same analysis in the whole brain (*SI Appendix, Table S3*).

We ensured independence between voxel selection of the clusters within the striatum during game play and the BOLD activation time course extracted

from the clusters. Striatal clusters exhibiting a significant Group  $\times$  Time Course activation were defined using each functional run separately. Given each run-wise striatal cluster, we extracted the average  $\beta$ -estimate time course from the remaining two functional runs (e.g., when the functional data from the first run were used to define a striatal cluster, the average  $\beta$ -estimate time course of the second and third functional runs of the videogame training was extracted from that cluster). We averaged the  $\beta$ -estimate time courses extracted from the three run-wise striatal clusters for each participant (Fig. 3*B*). The  $\beta$ -estimate peak around 7.5 s after the game event onset was used to quantify the extent of an individual's recruitment of the striatum during game play.

Note that across the three in-scanner videogame runs, the experimental and control groups did not exhibit significant differences in their behavioral game performances or in the game trial duration (*SI Appendix, Fig. S2*). Also, Group  $\times$  Time Course activation clusters within the striatum were quite consistent throughout the videogame training runs (*SI Appendix, Table S1*). These run-wise striatal clusters also served as seed regions in the functional connectivity analysis.

**Functional Connectivity Between the Striatum and Speech-Selective ROIs During Videogame Play.** We examined whether the striatal clusters involved in incidental sound category learning were functionally connected to the speech-selective regions, and whether the extent of the connectivity differed between the two participant groups. The active clusters within the striatum exhibiting Group  $\times$  Time Course effects served as seeds (*SI Appendix, Table S1*).

To quantify functional connectivity, we first constructed a new subject-level GLM for the videogame task to include additional nuisance regressors associated with baseline activity modulations during videogame training. We included the time-series data from two foci in the ventricles (62) and from one in the white matter. To remove effects of no interest, the modeled baseline activity was subtracted from the original time-series data, along with the modeled slow signal drift and head motion noted previously. Given this residual time-series data of each participant, we extracted average time-series data across the set of voxels within the striatal cluster involved during game play.

Using a similar analysis approach in extracting the activation (i.e.,  $\beta$ -estimate) time course of the striatal clusters, we also ensured the independence of the connectivity data and seed selection. In brief, regions within the striatum that exhibited Group  $\times$  Time Course in each run (*SI Appendix, Table S1*) served as seeds. The remaining residual time-series data (not used for defining the striatal seed region) were concatenated, and the average time series across the voxels in the corresponding seed was extracted. Given the seed region time series, we then computed voxel-wise temporal correlations for each participant. The resulting voxel-wise  $R$ -squared values were converted to Fisher's  $z$ -correlation coefficients. For each participant, we extracted the set of  $z$ -scored correlation coefficients across the voxels within individually defined speech-selective ROIs (Fig. 5, *Left*). This procedure was done for each of the run-wise striatal seed clusters (*SI Appendix, Table S1*). The average  $z$ -score correlation coefficient across the run-wise striatal clusters served as a measure of each participant's functional connectivity between the striatum and the speech-selective ROIs.

By using one-sample  $t$  tests against 0, we examined whether the striatum is functionally connected with the speech-selective ROI separately on each group. To contrast the extent of the connectivity between the two groups (experimental vs. control), we used an independent samples  $t$  test on the striatal connectivity to the speech-selective ROI. The same approach was used for the exploratory analysis on the striatal connectivity to visual and motor regions (*SI Appendix, Fig. S5*).

**Relationship Between Neural and Behavioral Modulations.** We performed a multiple linear regression analysis to examine whether the two participant groups exhibited a different relationship between the extent of the striatal modulations (i.e., striatal activation and its connectivity with speech-selective auditory ROIs) and individuals' behavioral performance. Individuals' striatal activation was quantified as the  $\beta$ -estimate peak extracted from the active clusters (Fig. 3). The average functional connectivity measure ( $z$ -transformed values) extracted from each speech-selective ROI served as individuals' striatal connectivity strength to auditory regions. With these striatal modulation measures, the multiple regression analyses were performed to predict the behavioral performances—namely, the behavioral posttest categorization of the novel exemplars of the onset-sweep categories, and the behavioral videogame performance (i.e., average highest level reached during the training). A significant Group  $\times$  Striatal Modulation (i.e., activation and connectivity) effect was followed up by a separate post hoc correlation analysis for each group of participants.

**ACKNOWLEDGMENTS.** The authors thank Mark Wheeler and Tyler Perrachione for helpful comments and discussions; Arava Kallai, Josh Tremel, and Corrine Durisko for their assistance with the statistical analyses; and Anthony Kelly for help running the experiment. This work was supported by training grants

(to S.-J.L.) from the NIH (T32GM081760 and 5T90DA022761-07) and the NSF (DGE0549352), grants (to L.L.H.) from the NIH (R01DC004674) and NSF (22166-1-1121357), and grants (to J.A.F.) from the NIH (R01HD060388) and NSF (SBE-0839229).

- Seger CA (2008) How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neurosci Biobehav Rev* 32:265–278.
- Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM (1998) A neuropsychological theory of multiple systems in category learning. *Psychol Rev* 105:442–481.
- Ashby FG, Ennis JM, Spiering BJ (2007) A neurobiological theory of automaticity in perceptual categorization. *Psychol Rev* 114:632–656.
- Seger CA, Cincotta CM (2005) The roles of the caudate nucleus in human classification learning. *J Neurosci* 25:2941–2951.
- Seger CA, Miller EK (2010) Category learning in the brain. *Annu Rev Neurosci* 33:203–219.
- Lopez-Paniagua D, Seger CA (2011) Interactions within and between corticostriatal loops during component processes of category learning. *J Cogn Neurosci* 23:3068–3083.
- Desai R, Liebenthal E, Waldron E, Binder JR (2008) Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci* 20:1174–1188.
- Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *J Neurosci* 29:5234–5239.
- Liebenthal E, et al. (2010) Specialization along the left superior temporal sulcus for auditory categorization. *Cereb Cortex* 20:2958–2970.
- Ley A, et al. (2012) Learning of new sound categories shapes neural response patterns in human auditory cortex. *J Neurosci* 32:13273–13280.
- Callan DE, et al. (2003) Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage* 19:113–124.
- Tricomi E, Delgado MR, McClelland BD, McClelland JL, Fiez JA (2006) Performance feedback drives caudate activation in a phonological learning task. *J Cogn Neurosci* 18:1029–1043.
- McClelland JL, Fiez JA, McClelland BD (2002) Teaching the /r-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiol Behav* 77:657–662.
- Goudbeek M, Swingle D, Smits R (2009) Supervised and unsupervised learning of multidimensional acoustic categories. *J Exp Psychol Hum Percept Perform* 35:1913–1933.
- Lim S-J, Fiez JA, Holt LL (2014) How may the basal ganglia contribute to auditory categorization and speech perception? *Front Neurosci* 8:230.
- Yi HG, Maddox WT, Mumford JA, Chandrasekaran B (2016) The role of corticostriatal systems in speech category learning. *Cereb Cortex* 26:1409–1420.
- Ashby FG, Ell SW (2001) The neurobiology of human category learning. *Trends Cogn Sci* 5:204–210.
- Ashby FG, Maddox WT (2005) Human category learning. *Annu Rev Psychol* 56:149–178.
- Seger CA (1994) Implicit learning. *Psychol Bull* 115:163–196.
- Foerde K, Shohamy D (2011) The role of the basal ganglia in learning and memory: Insight from Parkinson's disease. *Neurobiol Learn Mem* 96:624–636.
- Lim S-J, Holt LL (2011) Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cogn Sci* 35:1390–1405.
- Lim S-J, Lacerda F, Holt LL (2015) Discovering functional units in continuous speech. *J Exp Psychol Hum Percept Perform* 41:1139–1152.
- Wade T, Holt LL (2005) Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. *J Acoust Soc Am* 118:2618–2633.
- Liu R, Holt LL (2011) Neural changes associated with nonspeech auditory category learning parallel those of speech category acquisition. *J Cogn Neurosci* 23:683–698.
- Gabay Y, Holt LL (2015) Incidental learning of sound categories is impaired in developmental dyslexia. *Cortex* 73:131–143.
- Cincotta CM, Seger CA (2007) Dissociation between striatal regions while learning to categorize via feedback and via observation. *J Cogn Neurosci* 19:249–265.
- Nomura EM, et al. (2007) Neural correlates of rule-based and information-integration visual category learning. *Cereb Cortex* 17:37–43.
- Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. *Science* 273:1399–1402.
- Shohamy D, Myers CE, Kalanithi J, Gluck MA (2008) Basal ganglia and dopamine contributions to probabilistic category learning. *Neurosci Biobehav Rev* 32:219–236.
- Foerde K, Shohamy D (2011) Feedback timing modulates brain systems for learning in humans. *J Neurosci* 31:13157–13167.
- Seger CA, Spiering BJ (2011) A critical review of habit learning and the basal ganglia. *Front Syst Neurosci* 5:66.
- McClelland BD, Fiez JA, Protopoulos A, Conway M, McClelland JL (2002) Success and failure in teaching the [r-]l contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn Affect Behav Neurosci* 2:89–108.
- Yi H-G, Chandrasekaran B (2016) Auditory categories with separable decision boundaries are learned faster with full feedback than with minimal feedback. *J Acoust Soc Am* 140:1332–1335.
- Emberson LL, Liu R, Zevin JD (2013) Is statistical learning constrained by lower level perceptual organization? *Cognition* 128:82–102.
- Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. *Neuron* 41:281–292.
- Delgado MR, Stenger VA, Fiez JA (2004) Motivation-dependent responses in the human caudate nucleus. *Cereb Cortex* 14:1022–1030.
- Tremel JJ, Laurent PA, Wolk DA, Wheeler ME, Fiez JA (2016) Neural signatures of experience-based improvements in deterministic decision-making. *Behav Brain Res* 315:51–65.
- Shohamy D, et al. (2004) Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain* 127:851–859.
- Zheng T, Wilson CJ (2002) Corticostriatal combinatorics: The implications of corticostriatal axonal arborizations. *J Neurophysiol* 87:1007–1017.
- Flaherty AW, Graybiel AM (1994) Input-output organization of the sensorimotor striatum in the squirrel monkey. *J Neurosci* 14:599–610.
- Wilson CJ (1995) The contribution of cortical neurons to the firing pattern of striatal spiny neurons. *Models of Information Processing in the Basal Ganglia*, eds Houk JC, Davis JL, Beiser DG (MIT Press, Cambridge, MA), pp 29–50.
- Percheron G, Francois C, Yelnik J, Fenelon G, Talbi B (1994) The basal ganglia IV. *The Basal Ganglia Related Systems of Primates: Definition, Description and Informational Analysis*, eds Percheron G, McKenzie GM, Feger J (Plenum, New York), pp 3–20.
- Bar-Gad I, Morris G, Bergman H (2003) Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog Neurobiol* 71:439–473.
- Haruno M, et al. (2004) A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *J Neurosci* 24:1660–1665.
- Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol* 84:3072–3077.
- Elliott R, Newman JL, Longe OA, William Deakin JF (2004) Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *Neuroimage* 21:984–990.
- Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci* 29:2225–2232.
- Chandrasekaran B, Yi H-G, Maddox WT (2014) Dual-learning systems during speech category learning. *Psychon Bull Rev* 21:488–495.
- Ashby FG, O'Brien JB (2005) Category learning and multiple memory systems. *Trends Cogn Sci* 9:83–89.
- Seger CA, Peterson EJ (2013) Categorization = decision making + generalization. *Neurosci Biobehav Rev* 37:1187–1200.
- Seger CA, Peterson EJ, Cincotta CM, Lopez-Paniagua D, Anderson CW (2010) Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. *Neuroimage* 50:644–656.
- Ingalvalson EM, Holt LL, McClelland JL (2012) Can native Japanese listeners learn to differentiate /r-/l/ on the basis of F3 onset frequency? *Biling (Camb Engl)* 15:434–435.
- Iverson P, et al. (2003) A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87:B47–B57.
- Goldstein MH, King AP, West MJ (2003) Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc Natl Acad Sci USA* 100:8030–8035.
- Kuhl PK, Tsao F-M, Liu H-M (2003) Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci USA* 100:9096–9101.
- Gros-Louis J, West MJ, Goldstein MH, King AP (2006) Mothers provide differential feedback to infants' prelinguistic sounds. *Int J Behav Dev* 30:509–516.
- Seger CA (2013) The visual corticostriatal loop through the tail of the caudate: Circuitry and function. *Front Syst Neurosci* 7:104.
- Golestani N, Zatorre RJ (2004) Learning new sounds of speech: Reallocation of neural substrates. *Neuroimage* 21:494–506.
- Cox RW (1996) AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Talairach J, Tournoux P (1988) *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging* (Thieme Medical, New York).
- Delgado MR, Locke HM, Stenger VA, Fiez JA (2003) Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cogn Affect Behav Neurosci* 3:27–38.
- Windischberger C, et al. (2002) On the origin of respiratory artifacts in BOLD-EPI of the human brain. *Magn Reson Imaging* 20:575–582.