Neural systems involved in processing novel linguistic constructions and their visual referents

Matthew A. Johnson, Nicholas B. Turk-Browne, & Adele E. Goldberg

Department of Psychology
Princeton University, Peretsman-Scully Hall, Princeton, NJ, 08540

Please address correspondence to:
Adele E. Goldberg
Peretsman-Scully Hall
Princeton University
Princeton, NJ, 08544
609-258-8772
adele@princeton.edu

Running Title: Prediction from novel constructions

Abstract

In language, abstract phrasal patterns provide an important source of meaning, but little is known about whether or how such constructions are used to predict upcoming visual scenes. Findings from two fMRI studies indicate that initial exposure to a novel construction allows its semantics to be used for such predictions. Specifically, greater activity in the ventral striatum, a region sensitive to prediction errors, was linked to worse overall comprehension of a novel construction. Moreover, activity in occipital cortex was attenuated when a visual event could be inferred from a learned construction, which may reflect predictive coding of the event. These effects disappeared when predictions were unlikely: that is, when phrases provided no additional information about visual events. These findings support the idea that learners create and evaluate predictions about new instances during comprehension of novel linguistic constructions.

There is a growing recognition that prediction is important for language comprehension. We know that language is understood incrementally, with listeners anticipating upcoming words (Allopenna, Magnuson, & Tanenhaus, 1998; Arnold, Wasow, Losongco, & Ginstrom, 2000; Elman 2004; Gibson, Bergen & Piantadosi 2013; Kamide, Altmann, and Haywood 2003; Marslen-Wilson, 1973; Pickering & Garrod, 2013; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Traxler et al. 2000). For example, the N400 component in electroencephalography — an event-related potential related to semantic processing (King & Kutas, 1995; Kutas & Hillyard, 1984) — is reduced when words are predictable in a sentence. Prediction is possible because language involves probabilistic regularities that constrain our expectations and provide cues about upcoming words, phrases, and content (Bates & MacWhinney, 1987; Glenberg & Gallese, 2012; Goldberg, Casenhiser, & Sethuraman, 2005; Pickering & Garrod, 2004, 2007; Van Berkum et al. 2005). The ability to use word predictability to facilitate comprehension under noisy input conditions is positively correlated with sensitivity to transitional probabilities (Conway, Bauernschmidt, Huang, & Pisoni, 2010). Relatedly, sensitivity to sequential predictions correlates positively with the ability to successfully process complex, long-distance dependencies in natural language (Misyak, Christiansen, & Tomblin, 2010).

One type of probabilistic regularity is the association between certain abstract phrasal forms and particular abstract meanings. For example, the form "(NP_1) V NP_2 NP_3", as in "She mooped him something," is associated with an interpretation of 'transfer' (i.e., she transferred something to him by "mooping"). Although the terminology differs, all linguistic and psycholinguistic theories agree that these correspondences exist, and that they are a crucial part of one's knowledge, allowing speakers to creatively produce and understand utterances that they have never heard before (e.g., Christiansen & Chater, 1994; Goldberg, 1995, 2006; Kemmerer, 2006; Culicover & Jackendoff, 2005). We refer to these learned correspondences between form and function in what follows as grammatical *constructions*.

There has been considerable behavioral work investigating whether the semantics associated with familiar constructions such as the English transitive construction are comprehended by children at certain ages (e.g., Gertner, Fisher and Eisengart 2006; Hirsh-Pasek & Golinkoff 1996; Tomasello 2000). In addition, much computational work has been aimed at modeling the learning of grammatical constructions in the context of visual scenes (Dominey 2002; 2003; Dominey & Boucher 2005; Alishahi & Stevenson 2007).

At the same time, the majority of work using fMRI scanning methodology has focused on the processing of familiar or ungrammatical sentences, or on the learning or processing of words (e.g., Chow, Kaup, Raabe, & Greenlee, 2008; Hoen et al. 2006; Noordzij, Neggers, Ramsey & Postma, 2008; Xue, Chen, Jin, & Dong 2006). Other work on *artificial grammar learning* has focused specifically on the learning and recognition of formal regularities, for example, the ordering of words or letters (e.g., Grodzinsky & Friederici, 2006; Hoen et al. 2006; Lieberman, Chang, Chiao, Bookheimer, & Knowlton, 2004; Opitz & Friederici, 2004; Skosnik et al., 2002). Similarly, research on *statistical learning* has characterized the extraction of formal regularities from undifferentiated input (Abla, Katahira, & Okanoya, 2008; Turk-Browne, School, Chun, & Johnson, 2009; Karuza et al., 2013).

It is useful to distinguish between artificial grammar learning, which involves only formal regularities, and artificial *language* learning, which assigns interpretations to novel words and novel word orders. Our current focus is on the latter type of learning. There is a good deal of behavioral work on artificial language learning (e.g., Amato & MacDonald 2010; Braine et al. 1990; Kaschak & Glenberg, 2004; Kersten & Earles 2001; Wonnacott, Newport, & Tanenhaus,

2008; Wonnacott, Boyd, Thompson & Goldberg 2012). However, to date, there has not been much neuroscientific investigation of artificial language learning (beyond the level of novel word learning). This is likely due to the fact that learning several novel words, word orders, and meanings requires extensive exposure. Moreover, there have been no prior investigations, to our knowledge, of the neural processing of *novel constructions*, in which an unfamiliar abstract meaning is assigned to a novel formal pattern. That is, just as novel words can be assigned unfamiliar or familiar meanings (e.g., *blick* can be "a spiral shaped object" or "dog") (Bloom 1975; Clark 1995), so too can novel word order patterns can be assigned unfamiliar or familiar interpretations. The learning and processing of novel constructions—novel word orders paired with systematic, novel interpretations—provides a good case study of the task faced by language learners, since languages routinely pair formal patterns with interpretive functions in different ways (Croft 2001; Haspelmath 2008).

The present work uses neuroimaging to investigate how individual novel constructions are initially processed in the context of visual scenes. In particular, we assigned to various novel word orders novel abstract interpretations such as "X appears in location Y," "X approaches Y," or "X departs from location Y." Previous behavioral studies have shown that individual novel constructions can be learned relatively quickly by older children and adults when familiar nouns and novel verbs are used (Casenhiser & Goldberg 2005; Goldberg, Casenhiser, & Sethuraman, 2004). We take advantage of this rapid learning to investigate the neural activity evoked by novel constructions during initial exposure and immediately afterwards. Specifically, we investigate how the interpretations of novel linguistic constructions are used during comprehension in the context of visual scenes. We controlled for the visual stimuli and the processing of words with two *random* control conditions, as described below. The results of the present studies suggest that adult participants use nascent knowledge of grammatical constructions to predict the content of visual scenes.

Experiment 1

Because the processing of novel constructions in the brain had not previously been examined, we first conducted an exploratory fMRI study that probed the processes involved during initial exposure. This involved participants listening to tokens of a novel phrasal form and watching accompanying videos that conveyed abstract meaning. We created two novel constructions that differed in whether they conveyed predictive information about the accompanying visual scenes or not. In order to allow each participant to learn two distinct constructions, the constructions were also assigned distinct word orders and abstract interpretations as described below. For each construction, we evaluate the neural results by contrasting two conditions: *patterned*, in which the mapping of phrases onto videos was consistent, and *random*, in which the words were scrambled.

Methods

Participants. Eighteen right-handed native English speakers participated in exchange for monetary compensation. Data from two participants were discarded due to technical problems (their runs ended prematurely). The final sample of 16 participants (age 18-32; 9 female) is included in the analyses below. The study protocol was approved by the Princeton University IRB. Informed consent was provided by all participants.

4

Design. There were two phases: exposure and test. During the exposure phase, participants passively observed tokens of two novel constructions: an *appearance* construction and an *approach* construction.

The form of the appearance construction (Figure 1A) was "$V_{nonsense}$ $NP_{1(theme)}$ $NP_{2(location)}$". The meaning involved the theme (named by the first noun phrase, NP_1) appearing on or in the location (named by NP_2). Each instance of the construction was produced before the entity appeared on or in the named location. Therefore, knowledge of this construction would allow learners to predict that a particular entity would appear in or on a particular location. Each video opened with a different location (e.g., looking down at a table top), followed by the auditory narration 1-2 s later (e.g., "voopoes the bug the table"), and then the appearance of the character 1 s after the narration ended (e.g., the bug appearing on the table).

The form of the approach construction (Figure 1B) was "$NP_{1(agent)}$ $NP_{2(goal)}$ $V_{nonsense}$". The meaning involved the agent (named by NP_1) approaching the goal (named by NP_2). In all approach scenes, the agent was initially positioned to the left of another person (the goal). The narration began 1-2 s after this initial frame, and then the character on the left approached the character on the right 1 s after the narration ended, using various manners of motion. Although the timing matched the appearance construction, the auditory input in the approach construction did not carry additional predictive value about the visual event: the two people appeared before the narration started, and the person on the left always moved toward the person on the right. As a result, the appearance and approach constructions differed in the extent to which the narration was a helpful cue to predict what would happen in the video.
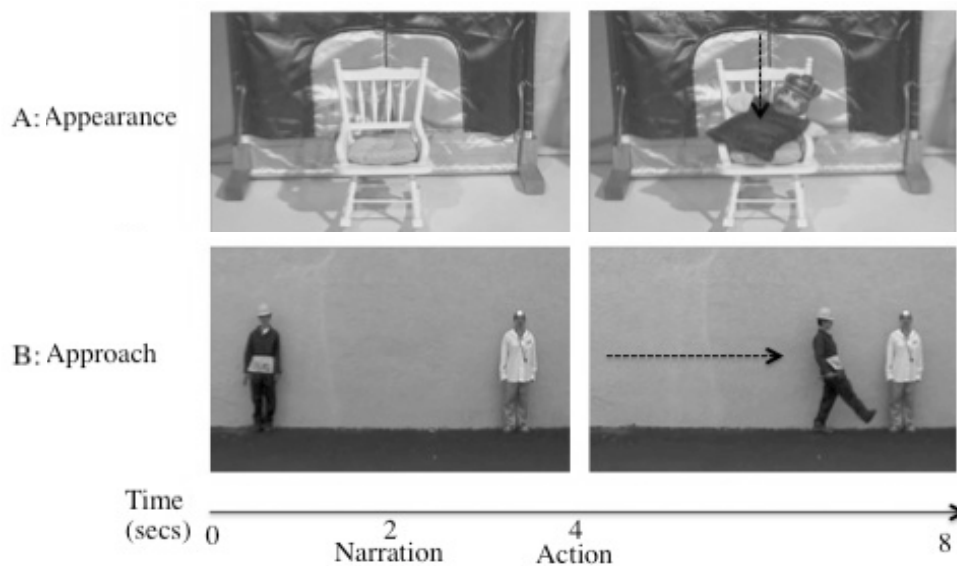
Figure 1: (A) Example of the appearance construction from Experiment 1 (e.g., "Moopoes the king the chair"). (B) Example of the approach construction from Experiment 1 (e.g., "The construction worker the doctor vakoes").

For each construction, there was a *patterned* condition and a *random* control condition. In the patterned condition, the pairing of the depicted actions and the linguistic narrations was consistent, enabling an abstract generalization to be learned. In the random condition, the same videos were viewed in the same sequence, but the order of words within each narration was scrambled. For example, instead of consistently hearing "NP_1 NP_2 $V_{nonsense}$" for approach scenes, participants were equally likely to hear "$V_{nonsense}$ NP_2 NP_1". Importantly, because the same videos and words were used in both the patterned and random conditions, neural differences between these conditions cannot be attributed to their visual and auditory features per se. Rather, the only thing that differed is that systematic language input in the patterned but not random condition allowed participants to extract form-meaning correspondences. Moreover, although the appearance and approach constructions differed in many ways, the comparison of patterned and random conditions within each construction allowed us to control for these differences and to isolate neural effects associated with learning the consistent pairing between narrations and actions.

Eight short (6-10 s) audio-visual tokens were witnessed for each condition with four fillers interspersed. The filler items served to impede participants from inferring a generalization from the visual scenes alone since the scenes presented were not uniformly appearance or approach. In order to maximize statistical power, the experiment was conducted within-subjects: each participant completed four scanning runs, one for each of four conditions (Table 1): appearance-patterned, appearance-random, approach-patterned, and approach-random.

	Appearance Runs		Approach Runs	
	Construction (8)	Filler (4)	Construction (8)	Filler (4)
Patterned	"V NP_1 NP_2" *NP_1 appears on/in NP_2*	"NP_1 V NP_2" *NP_1 does something to NP_2*	"NP_1 NP_2 V" *NP_1 approaches NP_2*	"NP_1 and NP_2 are V-ing" *NP_1 and NP_2 perform an intransitive action*
Random	Shuffled narrations over same videos		Shuffled narrations over same videos	

Table 1: The four exposure conditions in Experiment 1. The text in quotations refers to the narration that the participants heard and the text in italics refers to what happened in the video (NP = noun phrase; V = verb).

During the test phase, knowledge of the two constructions was assessed by determining how accurately participants could generalize beyond the exemplars witnessed in the exposure phase. Specifically, we tested whether participants successfully learned that the formal word order "NP_1 NP_2 V" has the abstract meaning that NP_1 approached NP_2 and that "V NP_1 NP_2" means that NP_1 appeared on or in NP_2. A forced-choice task was used in which participants heard a new instance of a witnessed phrasal form (i.e., with new nouns and a new nonsense verb) and had to choose which of two videos depicted the correct meaning.

Stimuli. All of the video clips used during exposure and test are available at http://www.mattajohnson.com/movies.html.

For the appearance construction, the videos included eight scenes of various objects appearing in or on various locations. The manners of motion varied across videos, including: falling from the sky, appearing out of nowhere, rolling into view, etc. The narration was "$V_{nonsense}$ $NP_{1(theme)}$ $NP_{2(location)}$", where NP_1 appeared on or in NP_2. Four filler clips were semantically transitive videos paired with transitive narrations (Table 1). Descriptions of the videos are provided in Table 2.

Type	Narration	Video description
Appearance cx exemplar	Voopos the king the chair.	A king puppet drops from in from above onto a chair.
Appearance cx exemplar	Zorpos the frog the box.	A frog drops in from above onto a box.
Appearance cx exemplar	Pippos the monster the cloth.	A monster puppet wriggles through a cloth into view.
Appearance cx exemplar	Habbos the queen the stage.	A queen rolls into view from offstage on the left.
Appearance cx exemplar	Kibbos the sun the sky.	A sun rises into the sky from offstage below.
Appearance cx exemplar	Vamos the bug the table.	A bug magically appears on the table.
Appearance cx exemplar	Napos the ball the room.	A ball rolls on stage from the left.
Appearance cx exemplar	Kubos the rabbit the hat.	A rabbit magically appears on the hat.
Filler	The frog poots the king.	A frog puppet pushes a king puppet.
Filler	The fireman zats the plane.	A fireman puppet turns a toy plane over.
Filler	The monster yebs the balloon.	A monster puppet shakes the balloon.
Filler	The clown goots the box.	A clown puppet puts his head in the box.

Table 2: The appearance construction exposure clips: eight exemplars of the appearance construction and four transitive fillers. All entities were puppets or toys. The form of the construction was V NP_1 NP_2, where "NP_1" appeared in or on "NP_2".

For the approach construction, the videos included eight scenes of a character on the left approaching a character on the right. The manners of motion varied across videos, including: swimming, jumping jacks, crab walk, marching, and walking. The narration was "NP_1 NP_2 $V_{nonsense}$", where NP_1 approached NP_2. Four filler clips were intransitive videos paired with conjoined intransitive narrations. Descriptions of the videos are provided in Table 3.

Type	Narration	Video description

Approach cx exemplar	The construction worker the doctor vakoes.	The construction worker twirls over to the doctor.
Approach cx exemplar	The construction worker the doctor tafoes.	The construction worker walks over to the doctor while moving her arms in a swimming motion.
Approach cx exemplar	The doctor the construction worker tafoes.	The doctor walks over to the construction worker while moving her arms in a swimming motion.
Approach cx exemplar	The doctor the construction worker moopos.	The doctor walks over to the construction worker while moving her legs in a kicking manner.
Approach cx exemplar	The construction worker the doctor sutos.	The construction worker walks over to the doctor while dragging one foot.
Approach cx exemplar	The doctor the construction worker vakoes.	The doctor twirls over to the construction worker.
Approach cx exemplar	The construction worker the doctor moopos.	The construction worker walks over to the construction worker while moving her legs in a kicking manner.
Approach cx exemplar	The doctor the construction worker sutos.	The doctor walks over to the construction worker while dragging one foot.
Filler	The construction worker and the doctor are koobing.	The construction worker and the doctor are waving.
Filler	The doctor and the construction worker are gooding	The doctor and construction worker are rubbing their bellies.
Filler	The construction worker and the doctor are pimming.	The construction worker and the doctor are tapping their heads.
Filler	The doctor and the construction worker are chibbing.	The doctor and the construction worker are bowing.

Table 3: The approach construction exposure clips: eight exemplars of the appearance construction and four intransitive fillers. The construction worker and doctor were humans in costume. The form of the construction was $NP_1\ NP_2\ V$, where "NP_1" approaches "NP_2".

These two constructions were chosen because previous behavioral work indicated that they were easily learnable with brief exposure (Goldberg, Casenhiser, & Sethuraman, 2004; Boyd, Gottschalk, & Goldberg, 2009). Importantly, they also differed in the extent to which the narration was needed to predict what would happen in the video.

The test trials consisted of one narration and two simultaneous videos. One video was a novel target exemplar that was congruent with exposure and the other was a foil that was incongruent. Foil videos included the same two arguments as the target, but instead of appearance or approach, the two arguments performed either a semantically transitive action or a

conjoined intransitive action. The test trials for the appearance and approach constructions are described in Tables 4 and 5, respectively.

Narration	Simultaneous test alternatives (2AFC, R/L counterbalanced)	
	Appearance scene	**Foil scene**
Grebos the clown the box.	A clown pops out of a box.	A clown raises a box using a seesaw.
Bakos the flower the grass.	A flower grows out of the ground.	A flower grows taller (in view).
Veemos the king the bed.	The king flies in from offstage onto the bed.	The king jumps over the bed (onscreen).
Bupping the man the chair.	The man flies out of a cannon onto the chair.	A man shoots a gun at the chair.
Fippos the pig the ramp.	A pig rolls onto the ramp from off screen.	A pig rolls up and down on the ramp (on screen).
Nibbos the sailor the pond.	A sailor sails onto the screen.	A sailor sails around on screen.
The princess jorps the flower (foil).	A flower comes to appear in the princess' hand.	The princess bends down around the flower.
The wizard fims the bug (foil).	A bug flies in near the wizard's head.	The wizard bends over the bug.

Table 4: Appearance construction test items: new instances of the construction and fillers. All entities are puppets or toys.

Narration	Simultaneous test alternatives (2AFC, R/L counterbalanced)	
	Approach scene	**Foil scene**
The business person the boxer durfos.	The business person approaches the boxer on all fours.	The business person and the boxer are bending their knees simultaneously.
The angel the judge gitos.	The angel jumping jacks her way over to the judge.	The angel and the judge look right and left simultaneously.
The construction worker the doctor moopoes.	The construction worker marches over to the doctor.	The doctor marches over to the construction worker.
The doctor the construction worker sootoes.	The doctor approaches the construction worker, dragging her foot.	The construction worker approaches the doctor, dragging her foot.
The construction worker and the doctor are yurting (foil)	The construction worker jogs over to the doctor.	The construction worker and the doctor are kicking their leg forward simultaneously.

The doctor and the construction worker are shumming (foil)	The doctor approaches the construction worker by hopping.	The doctor and the construction worker are looking up to the sky and down again simultaneously.
The grandma and the cowgirl are fopping (foil)	The cowgirl crabwalks over to the grandma.	The grandma and the cowgirl are moving their hands up and down.
The basketball player and the princess are kadding. (foil)	The princess approaches the basketball player by hopping over on one foot.	The basketball player and the princess are clapping simultaneously.

Table 5: Approach construction test items: new instances of the construction and fillers. All entities were humans in costume.

     Procedure. During the exposure phase, participants were instructed to simply "watch the videos and pay attention to the language that describes them." No further explicit instruction was provided. Participants were given a button box upon entering the scanner, and were told that it would be used in a second study that was separate from the video-viewing task. The order of the 12 videos for each construction was randomized for each participant, but this order was held constant for the patterned and random runs of each construction. The order of runs (conditions) was counterbalanced across participants, except that the patterned and random conditions of a given construction never occurred back-to-back; this was done to avoid direct comparison of patterned and random versions of the same construction. By counterbalancing the serial position of each condition, neural effects that are reliable across participants cannot be attributed to the order of conditions or to how attention fluctuated during the study. Participants could take a brief break between runs, and the experimenter started each new run with "OK, here comes the next one." This provided a subtle cue that the next set of stimuli was distinct from previous set.

     During each trial of the test phase, participants listened to the narration and viewed the videos, and were then asked: "Did the narration match the video on the left or on the right?" They signaled their response with the button box. The response window lasted 10 s, after which the screen turned black until the next trial commenced. There were eight trials in each test run, spaced by 30 s intervals, and one test run for each construction.

     fMRI acquisition. Scanning was performed on a Siemens 3T Allegra head-only scanner using a Nova Medical NM-011 Head Transmit Coil with receive-only array system. Functional sequences were acquired using a T2*-weighted EPI sequence (echo time = 28 ms, repetition time = 2000 ms, flip angle = 71°, matrix = 64 x 64). Twenty-six interleaved oblique axial slices were aligned parallel to the anterior-/posterior-commissure, and whole-brain coverage was obtained at a resolution of 3.5 mm isotropic voxels. For alignment, two T1-weighted structural images were acquired: a coplanar FLASH image and a high-resolution MPRAGE image.

     fMRI preprocessing. The first four volumes (8 s) were discarded from each run to allow for T1 equilibration. Data were preprocessed in FSL (www.fmrib.ox.ac.uk/fsl), including: corrections for head motion and slice-acquisition time, high-pass temporal filtering (100 s period cut-off), and spatial smoothing (5 mm kernel). Functional runs were registered to the FLASH image, the MPRAGE image, and the standard Montreal Neurological Institute (MNI) brain (and interpolated to 2 mm voxels).

     fMRI analysis. Each run was fit using a general linear model (GLM) that contained one regressor modeling the average blood-oxygen level-dependent (BOLD) response across videos in

that run. The baseline in each run consisted of the rest periods between videos. Each video was modeled with a boxcar function convolved with a canonical hemodynamic response function. Since learning the constructions required noticing shared structure across exemplars, we excluded the first two videos in each run from analysis. To compare across conditions, we contrasted parameter estimates obtained in the first-level GLMs in a higher-level GLM treating subject as a random effect. Specifically, we contrasted patterned vs. random separately for each construction, and also collapsing across constructions. A pre-threshold mask of brain voxels was generated by averaging images all conditions and subjects. The voxel-wise statistical maps from the higher-level GLM were thresholded at $p < 0.05$ corrected, based on a cluster-mass correction with an initial cluster-forming threshold of $z = 2.3$.

To examine brain-behavior relationships, we used individual behavioral test performance (overall and within each construction) as a regressor. We mean-centered the behavioral scores across participants and then fit new higher-level GLMs where the parameter estimate reflecting the BOLD response for each participant and condition was weighted by the normalized behavioral score. The resulting maps were thresholded in the same manner as above. The random runs served as a control: no behavioral correlation was expected because no learning relevant to the test trials was possible.

Results

Test performance confirmed that participants learned both novel constructions (Figure 2). Across participants, mean accuracy was 73% (SD = 17%) for the appearance construction and 80% (SD = 23%) for the approach construction. These levels were greater than chance (vs. 50%: appearance, $t(15) = 5.51$, $p < .001$; approach, $t(15) = 5.18$, $p < .001$), and they did not differ significantly from each other ($t(15) = 1.06$, $p = .30$).
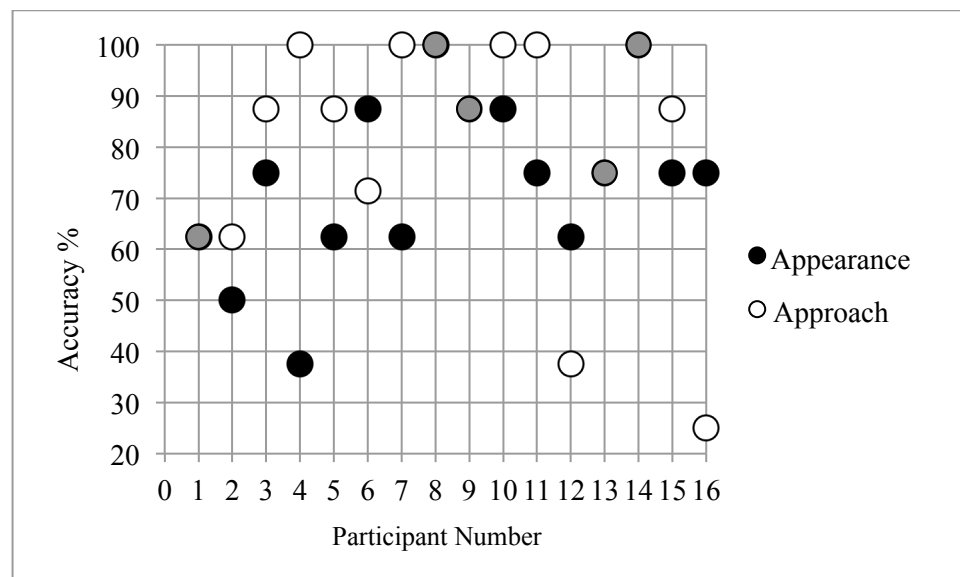
Figure 2: The distribution of test performance across participants in Experiment 1, for approach (white) and appearance construction (black); identical performance indicated in gray (chance = 50%).

11

Because of the considerable variability in behavioral performance, we began the neural analyses by examining brain-behavior relationships. Specifically, we correlated mean BOLD activation during the exposure phase with final test performance across participants using a covariate analysis. In the patterned condition, no regions exhibited a positive correlation with behavior. However, as shown in Figure 3A, we observed a robust negative correlation — i.e., lower activation during exposure predicted better test performance — centered on the ventral striatum ($p < .05$ corrected; peak MNI x, y, z coordinates: -20, 12, -8); this cluster also extended into the caudate, putamen, and midbrain. We further examined this inverse relationship with behavior in the ventral striatum for each construction. The effect was present for the appearance construction (Figure 3B), but not for the approach construction (Figure 3C). In the control (random) condition, no regions showed a significant positive or negative correlation with performance.

A: Overall Patterned   B: Patterned Appearance   C: Patterned Approach
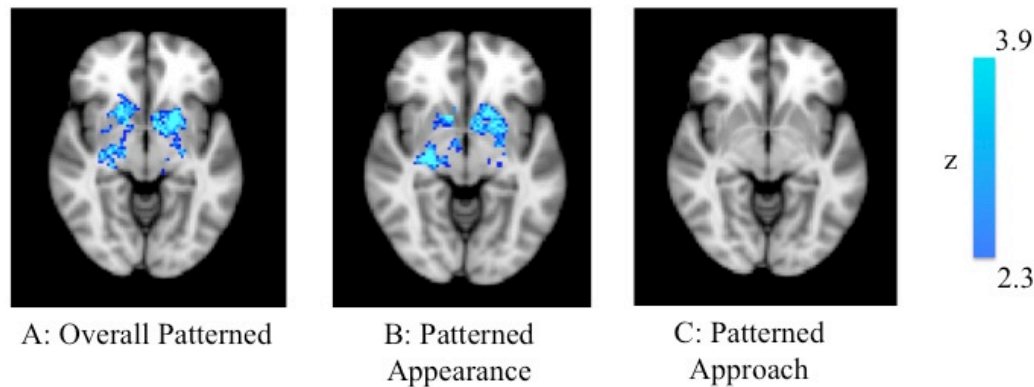
Figure 3: Correlation of test performance in Experiment 1 with activation in the patterned condition. (A) Collapsing over constructions, bilateral ventral striatum was negatively correlated with behavior ($p < .05$ corrected). This effect was robust for (B) the appearance construction, but was absent for (C) the approach construction.

We next examined neural responses during the exposure phase irrespective of behavior. No regions showed patterned > random collapsed across constructions. The medial prefrontal cortex (3, 48, 24) and left supramarginal gyrus (-42, -52, 32) showed random > patterned overall, although not for either construction individually. Instead, left occipital cortex (-22, -84, 8) showed random > patterned for the appearance construction (Figure 4); no regions were obtained for the approach construction. A summary of the results is provided in Table 6.

Analysis	Construction	Region	Peak	Extent	Significance
Negative correlation of patterned exposure with test behavior	Collapsed	Ventral striatum (bilateral)	-20, 12, -8	606	*p* < .001
Negative correlation of patterned exposure with test behavior	Collapsed	Midbrain	-2, -18, -14	573	*p* = .004
Negative correlation of patterned exposure with test behavior	Appearance	Ventral striatum (bilateral)	10, 12, 0	2336	*p* < .001
Random > patterned	Collapsed	Medial prefrontal cortex	3, 48, 24	389	*p* = .015
Random > patterned	Collapsed	Supramarginal gyrus (left)	-42, -52, 32	360	*p* = .024
Random > patterned	Appearance	Occipital cortex (left)	-22, -84, 8	615	*p* < .001

Table 6: Summary of whole-brain results from Experiment 1. Coordinates in MNI space (x, y, z). Extent refers to the number of voxels per cluster. Significance is corrected for multiple comparisons based on cluster mass.

Discussion

The appearance construction but not the approach construction provided an opportunity to predict upcoming visual events on the basis of the narrations. Specifically, in the appearance construction, the linguistic input identified the type of entity that would appear and the location in which the entity would appear. On the other hand, in the approach construction, witnessing the initial visual frame was sufficient to predict which onscreen entity would move, and in which direction (the character on the left always approached the character on the right). Although participants might have attempted to predict which novel verb would be used on the basis of the manner of motion involved, such prediction could not be expected to correlate with accuracy at test, since the test items and their foils involved new novel verbs.

We can tentatively interpret the neural results for the patterned appearance condition in light of the predictive aspect of this construction, for two reasons. First, the inverse correlation between behavioral performance and activity in the ventral striatum can be interpreted in terms of prediction error. In particular, discrepancies between predicted and actual outcomes are known to elicit responses in the ventral striatum (Niv & Schoenbaum, 2008; O'Doherty et al.,

2004; Pagnoni, Zink, Montague, & Berns, 2002). Since the narration of the appearance construction allowed participants to anticipate how the scene would unfold, successful learning of the construction may have reduced uncertainty and attenuated striatal responses.[1]

The inverse relationship between striatal activity during exposure and behavioral accuracy at test was not observed for the approach construction. We interpret this as reflecting the fact that learning this construction did not provide an advantage in predicting how the visual scenes would unfold. Thus, although participants may well have been trying to predict upcoming events in all conditions — including the random conditions in which there was no systematic relationship between the linguistic forms and the visual scenes — the only case in which more accurate prediction during exposure (as evidenced by reduced responses in ventral striatum) could be expected to correlate with knowledge of the construction (as evidenced by accuracy at test) would be the patterned appearance run.

Appearance:
Random > Patterned
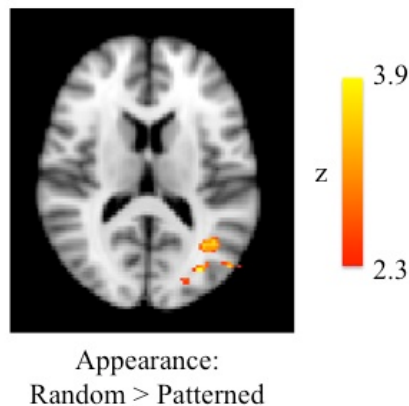
Figure 4: Random > patterned in the left occipital cortex during exposure to the appearance construction in Experiment 1 ($p < .05$ corrected).

The second piece of evidence for an interpretation in terms of prediction was the reduced activation in occipital cortex for the patterned condition compared to the random condition in the appearance construction. This occipital deactivation can be related to predictive coding: generating expectations about impending stimuli can facilitate visual processing and attenuate the neural activity (Kok, Jehee, & de Lange, 2012; Meyer & Olson, 2011; Summerfield & Egner, 2009). Again, the patterned and random appearance conditions were identical in terms of visual

---

[1] Note that the negative relationship between ventral striatum activity and test performance may initially seem counterintuitive, as the size of a prediction error is related to the amount of ensuing learning (e.g., Rescorla & Wagner 1972). However, remember that we examined brain activity only after the first two tokens of each construction had been witnessed during exposure, which allowed initial learning to take place. We thus interpret activity in ventral striatum for later tokens as indicative of continued errors — i.e., as evidence of insufficient knowledge of the construction for accurate predictions to be generated. The negative relationship between this activity and test performance is consistent with such an interpretation.

input, suggesting that the consistent pairing of the narration allowed participants to generate more specific expectations about the video from the narration. This was especially important for the appearance condition since the narration allowed listeners to expect that a particular type of entity would appear in a particular location.

It is perhaps surprising that no part of the language-relevant network (e.g., Grodzinsky & Friederici, 2006; Kaan & Swaab, 2002; Osterhout, Kim, & Kuperberg, 2007; Silbert et al. 2014; Stowe,Haverkort, & Zwarts, 2005) displayed a distinctive response to the abstract construction in the patterned condition when compared with the random condition. This may reflect the fact that meaningful words were used in both patterned and random runs, and words and constructions are processed in similar brain regions (Fedorenko, Nieto-Castanon & Kanwisher 2012). In any event, the lack of differential activation in the language-relevant network suggests that the present results may not capture how constructions are neurally represented, and therefore may not be useful for adjudicating between neurolinguistic models (e.g., Arbib & Lee 2008; Dominey, Hoen, & Inui 2006; Friederici, Opitz, & von Cramon 2000; Pulvermüller, Cappelle & Shtyrov 2013; Ullman 2004). At the same time, the present results provide intriguing new evidence for domain-general predictive processes in language processing.

Experiment 2

If prediction is driving both the negative behavioral correlation in the ventral striatum and the occipital deactivation during construction learning, we should be able to obtain similar findings in an experiment specifically designed to identify a predictive advantage. As Experiment 1 was exploratory, the appearance and approach constructions were not designed to differ only in the predictability of the visual scenes. In fact, they differed on a number of other dimensions as well, including: puppets and toys (appearance) vs. human actors in costume (approach); more than two distinct entities (appearance) vs. exactly two actors (approach); and verb initial (appearance) vs. verb final (approach). Therefore, we developed a new *departure* construction that matched the approach construction in terms of these other dimensions (verb final, two human actors in costume) but matched the appearance construction only along the critical dimension — carrying predictive information about the visual scene. In this way, multiple aspects of the two predictive constructions in Experiments 1 and 2 varied, while the fact that both served to predict the upcoming visual scenes was held constant. The details of these three constructions are provided in Table 7.

Construction	Neural findings	Predictability	Phrase	Entities
Appearance cx (Experiment 1)	Ventral striatum and occipital involvement for patterned relative to random conditions	Narration served to help predict visual scene	V NP NP	Various puppets and toys
Approach cx (Experiment 1)		Visual scene predictable regardless of narration	NP NP V	Two humans in costumes
Departure cx (Experiment 2)	Hypothesized ventral striatum and occipital involvement for congruent relative to incongruent conditions	Narration served to help predict visual scene	NP NP V	Two humans in costumes

Table 7: Similarities and differences between constructions used in Experiments 1 and 2.

Recall that in Experiment 1, we found an inverse relationship between activity in the ventral striatum during exposure and behavioral accuracy at test. We inferred from this that the ventral striatum served as an index of prediction error and that greater errors during exposure (reflecting worse learning) led to worse test performance. In order to probe this brain-behavior relationship more directly, Experiment 2 examines the link between striatal activity and behavioral performance within the test phase. We compared neural responses for exemplars that were either *congruent* or *incongruent* with a learned construction. For each exemplar, participants judged whether the video was narrated correctly (congruent) or not (incongruent).

Although Experiment 1 was exploratory, it helped us generate three specific hypotheses for Experiment 2. First, insofar as the ventral striatum tracks prediction error during the processing of linguistic constructions, then lower activity during congruent test trials (reflecting fulfilled expectations) should be associated with greater behavioral accuracy (i.e., more likely to report that yes, the video was narrated correctly). Second, and relatedly, *higher* activity during incongruent test trials (reflecting violated expectations) should also be associated with greater behavioral accuracy (i.e., more likely to report that no, the video was not narrated correctly). Third, and finally, insofar as the left occipital cortex deactivates for predictable visual events, then it should respond less to congruent vs. incongruent trials. To be as hypothesis-driven as possible, we tested these hypotheses in Experiment 2 within regions of interest (ROIs) defined based on the clusters in ventral striatum and left occipital cortex obtained from Experiment 1.

Note that we focus analyses on the test phase because we believe that this is the most sensitive way to test our hypotheses about prediction. This was not possible in Experiment 1 because we did not collect fMRI data during the test phase, and so we settled for relating exposure fMRI data to test behavioral data. However, prediction during the exposure phase was necessarily diluted by the learning in this phase needed to be able to generate predictions in the first place. The test phase thus provides a clearer measure of prediction because any learning that

was going to happen had occurred by this point. In addition, we doubled the amount of exposure in Experiment 2 to provide as much opportunity as possible for learning to occur.

Methods

Participants. A distinct group of twenty right-handed native English speakers participated in this experiment. Data from four participants were discarded due to technical problems (two because runs ended prematurely, and two because behavioral responses at test were accidentally not recorded). Data from the remaining 16 participants (age 19-32; 8 female) are reported below.

Design. A novel *departure* construction (Figure 5A) was created for this experiment with the form "NP_1 NP_2 $V_{nonsense}$". The meaning always involved the argument designated by NP_2 moving away from the argument designated by NP_1. This construction was designed to mirror the appearance construction in providing a predictive cue: until the narration, it was unclear which of the two characters would depart the scene. At the same time, this construction also employed the same phrasal form and number and type of actors as the non-predictive approach construction.

A: Departure construction
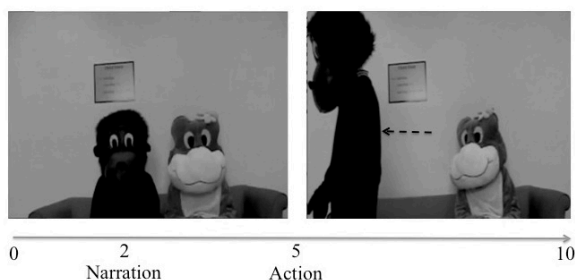
Figure 5: Example of the departure construction from Experiment 2 (e.g., "The frog the gorilla vakoes").

The test phase consisted of new exemplars that had not been witnessed during exposure, to assess learning of the abstract construction. Specifically, we tested whether participants successfully learned that the formal word order "NP_1 NP_2 V" has the abstract meaning that NP_2 departed from NP_1. Unlike Experiment 1 in which each test trial contained one narration and two videos (one congruent and one incongruent), the test trials in Experiment 2 consisted of one narration and one video (either congruent or incongruent).

Stimuli. Each of the 12 exposure videos was a scene with two humans, in gorilla and frog costumes, initially standing or sitting side-by-side, followed by one of the characters leaving. The videos lasted 12 s in duration, with the narration beginning after 1 s and finishing after 3-4 s, and the action beginning after 5 s. The manners of motion varied across videos, including: pushing away on a chair, walking away through a door, magically disappearing, etc. The narration was "NP_1 NP_2 $V_{nonsense}$", where NP_2 moved away from NP_1. Descriptions of the videos are provided in Table 8.

Narration	Video description
The gorilla the frog moopos.	The frog gets up and walks away from the gorilla to the right.

The frog the gorilla veemos.	The frog and gorilla are sitting and the gorilla magically disappears.
The gorilla the frog doopos.	The frog and gorilla are sitting and the frog gets up and walks away to the right.
The frog the gorilla vakos.	The frog and gorilla are sitting and the gorilla gets up and walks away to left.
The gorilla the frog packos.	The frog and gorilla are standing, and the frog leaves through a door.
The frog the gorilla sapos.	The gorilla and frog are standing and the gorilla magically disappears.
The gorilla the frog sapos.	The gorilla and frog are standing and the frog magically disappears.
The frog the gorilla pimmos.	The frog and the gorilla are standing and the gorilla leaves through a door on the left.
The gorilla the frog nulos.	The gorilla and the frog are sitting at a desk and the frog gets up and walks away.
The gorilla the frog jiffos.	The gorilla and the frog are sitting at a desk and the frog moves away by pushing his chair back.
The frog the gorilla zavoes.	The gorilla and the frog are sitting at a desk and the gorilla gets up and walks off screen.
The frog the gorilla jiffos.	The gorilla and the frog are sitting at a desk and the gorilla moves away by pushing his chair back.

Table 8: Exemplars of the departure construction used during exposure in Experiment 2. The frog and gorilla were humans in costume. The form of the construction was $NP_1\ NP_2\ V$, where "NP_2" moves away from "NP_1".

The test videos were similar to those witnessed during exposure, but required generalization because they involved new scenes of departure with new novel verbs and new manners of motion. There were two versions of each of the eight videos. The two versions differed only in terms of whether the accompanying narration was congruent (consistent with what was experienced in the exposure phase) or incongruent (inconsistent with exposure). In particular, NP_2 moved away from NP_1 on congruent trials, and NP_1 moved away from NP_2 on incongruent trials. Each participant saw only one version, counterbalanced such that each version was seen by half of the participants. The test trials are described in Table 9.

Narration	Scene 1	Scene 2
The dog the bear packos.	The dog and bear are sitting and the dog gets up and walks away.	The dog and bear are sitting and the bear gets up and walks away.
The penguin the chicken moolos.	The penguin steps away from the chicken.	The chicken steps away from the penguin.
The bear the dog jiffos.	The bear gets up from computer desk and walks away.	The dog gets up from computer desk and walks away.

18

The penguin the chicken pinkos.	The penguin walks away and sits down.	The chicken walks away through a door.
The dog the bear kafos	The dog magically disappears.	The bear magically disappears.
The bear the dog matos.	The bear walks away from the dog.	The dog walks away from the bear.
The penguin the chicken sabbos.	The penguin walks through a door away from the chicken.	The chicken walks through a door away from the penguin.
The chicken the penguin zoipos.	The chicken walks up the stairs and away from the penguin.	The penguin walks up the stairs and away from the chicken.

Table 9: Test items for the departure construction. Each narration was played with both of the scenes on different test trials. All entities were humans in costume. Half of the scenes are congruent and half are incongruent on each list.

Procedure. The exposure phase was similar to Experiment 1, except that there were two fMRI runs for the departure condition, which each contained all 12 exemplars. The test trials occurred during another fMRI run. After each trial, participants were asked: "Was the video narrated correctly?" Within all runs, exemplars were separated by a rest period of 10 s.

In addition to the construction exposure and test runs, this fMRI session also included interleaved runs from a separate word learning experiment. We do not report that experiment in this paper because its design was not comparable to the present study, nor was it pertinent to our hypotheses.

fMRI analysis. The fMRI acquisition and preprocessing were identical to Experiment 1, except that no additional volumes were discarded at the beginning of the run beyond those acquired and discarded by the scanner. The test run for each participant was fit using a GLM with two regressors that modeled the BOLD response for congruent and incongruent trials, respectively. The baseline was the rest periods between trials. We also performed a follow-up GLM analysis that separated congruent and incongruent trials based on correct vs. incorrect behavioral responses (hits vs. misses and correct rejections vs. false alarms, respectively). To compare conditions, we contrasted parameter estimates in higher-level GLMs, treating subject as a random effect in paired t-tests. The resulting voxel-wise statistical maps were thresholded in the same way as Experiment 1.

The primary analysis for Experiment 2 involved defining functional ROIs based on Experiment 1. For the ventral striatum, we generated a binary mask from the corrected map for the inverse correlation with behavioral performance in the appearance construction. For left occipital cortex, we generated a binary mask from the main effect of random > patterned in the appearance construction. These ROIs were then transformed into each subject's brain space. The BOLD response in each ROI was calculated by averaging the parameter estimates for a condition in the GLMs over all voxels in the ROI.

Results

Participants learned the departure construction well (Figure 6). The average behavioral test accuracy across participants was 77.5% (SD = 23%), which was above chance (vs. 50%: $t(15) = 4.85, p < .001$).
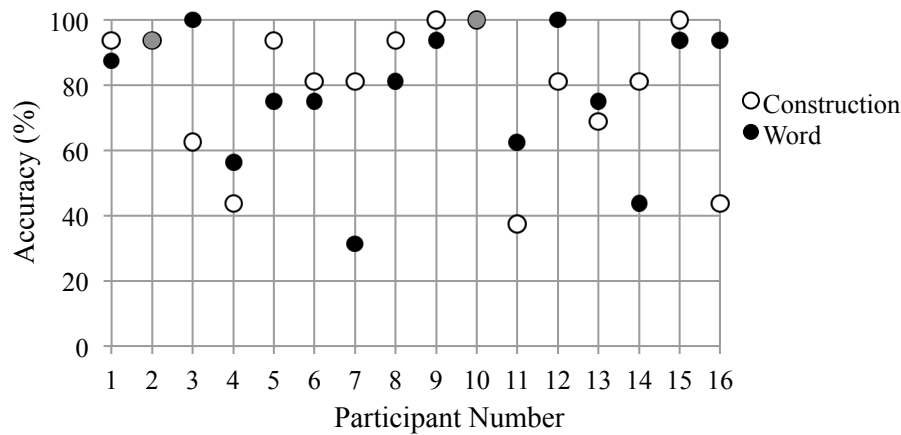

Figure 6: The distribution of test performance across participants in Experiment 2 (chance = 50%).

Our first hypothesis was that there would be less activity in the ventral striatum when participants could accurately predict what would happen in the video from the narration. We examined this in the test phase, when we were confident that learning had occurred, and for congruent trials, in which the video was consistent with the narration. Otherwise, we used the same approach as Experiment 1 — correlating brain activity with behavioral performance across participants. Consistent with our hypothesis, there was a negative correlation between activity in the ventral striatum ROI and test accuracy for the congruent trials ($r = -.47, p = .03$).

Our second hypothesis was that this relationship would flip for incongruent trials, with more activity in the ventral striatum associated with better behavioral performance. Namely, on trials where the video was inconsistent with the narration, greater activity would reflect a violation of the learned expectation. Indeed, there was a positive correlation between activity in the ventral striatum ROI and test accuracy on incongruent trials ($r = .46, p = .04$).

Our third hypothesis concerned the left occipital cortex, specifically that there would be less activity when participants could accurately predict what would happen in the video. We again examined this in the test phase after learning had occurred, and took the same analysis approach as in Experiment 1 of looking at overall differences in activity between predictable and unpredictable conditions. Partly consistent with the hypothesis, there was a numerical trend toward lower activity in the left occipital ROI for congruent vs. incongruent trials ($t(15) = 1.75, p = .10$). Upon further reflection, we realized that the overall activity in congruent and incongruent trials was not the most precise way of capturing predictable vs. unpredictable events. In particular, behavior provided additional information: congruent trials reflected confirmed predictions only when participants correctly identified the video as consistent with the narration (hits in signal detection terms), and incongruent trials reflected violated predictions only when participants correctly identified the video as inconsistent with the narration (correct rejections). Therefore, we restricted analyses to these correct trials (77.5% of all trials; 73% average hit rate for congruent trials, 81% average correct rejection rate) and excluded incorrect trials (i.e., misses

and false alarms) in which the prediction status was unclear. After doing so, we found the hypothesized deactivation of the left occipital ROI for congruent hits vs. incongruent correct rejections ($t(15) = 3.35$, $p < .005$).

Left Occipital Cortex
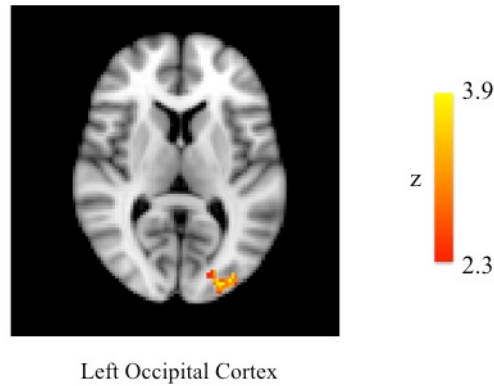
Figure 7: Experiment 2. At test, left occipital cortex showed deactivation for congruent vs. incongruent trials ($p < .05$ corrected).

In addition to the primary ROI analyses, we also conducted exploratory whole-brain analyses for the sake of completeness. A region of left occipital cortex (-20, -94, 8) near to the left occipital ROI showed an effect of incongruent > congruent ($p < .05$ corrected); this cluster also extended into the precuneus (-2, -76, 44). No regions showed an effect of congruent > incongruent.

Discussion

Experiment 2 provided converging evidence for the neural basis of the processing of linguistic constructions. The two most analogous findings to Experiment 1 — an inverse relationship between activity in the ventral striatum and behavior on congruent trials, and the relative deactivation of occipital cortex for correct congruent vs. incongruent trials — both suggest that constructions can be leveraged to predict upcoming visual events.

General Discussion

We started in Experiment 1 with an exploratory investigation of how linguistic constructions are processed in the brain. Based on the findings, we designed Experiment 2 to more directly test hypotheses about predictive processes induced by recently learned constructions. The two key brain systems highlighted in both studies, the ventral striatum and occipital cortex, have been studied extensively in the context of related and different mental processes. Below we selectively review aspects of these literatures that seem relevant for interpreting our findings.

Ventral Striatum

In Experiment 1, whole-brain analyses revealed an inverse relationship between activity in the ventral striatum during the exposure phase and subsequent behavioral performance in the test phase. In Experiment 2, this relationship was probed more directly with an ROI analysis

during the test phase. Once again, results yielded an inverse relationship between striatal activity and performance for congruent test items. This link was further strengthened by a dissociation in Experiment 2, with an opposite positive relationship for incongruent test items.

We interpret these results as reflecting prediction error in the ventral striatum. That is, the extent to which construction learning has occurred (as measured behaviorally) determines the strength with which constructions can be used to generate expectations. For participants who learned well, hearing the narration allowed them to generate a strong expectation of what would transpire visually. When this expectation was confirmed, either by a familiar exemplar (Experiment 1) or a new congruent exemplar (Experiment 2), the ventral striatum was less responsive. When this expectation was violated, such as by an incongruent exemplar, the ventral striatum responded more strongly. In contrast, for participants who did not learn well, they could at best generate a weak expectation, such that familiar and congruent exemplars were somewhat surprising; because the expectation was weak, incongruent exemplars were less surprising.

Although the ventral striatum is typically linked to reward prediction and not the sort of "stimulus" prediction of the videos that we infer (e.g., Niv & Schoenbaum, 2008; Schultz, Dayan, & Montague, 1997), there is some evidence that midbrain dopaminergic neurons that project to the ventral striatum are involved in the processing of salient stimuli (Ungless, 2004) and the more general prediction of information (Bromberg-Martin & Hikosaka, 2009; Redgrave & Gurney, 2006). For example, when given a choice about whether or not to receive information about an upcoming water reward, monkeys consistently seek out this information, even when they know that the size of the reward is independent of their choice. Not only do midbrain dopaminergic neurons respond to cues that predict water rewards, but also to such offers of information (Bromberg-Martin & Hikosaka, 2009). In this way, dopamine can signal the expectation of information in a more general sense.

Determining whether the ventral striatum is involved in predicting non-rewarding stimuli awaits further investigation. Although extrinsic rewards were not used in our studies, one alternative is that the involvement of the ventral striatum reflects *intrinsic* rewards (e.g., Botvinick, Niv, & Barto, 2009), possibly arising from successful learning or prediction itself. Regardless, our findings are consistent with the idea that the ventral striatum can support more general types of learning beyond classical error-driven learning of external reward contingencies.

Occipital deactivation

In Experiment 1, left occipital cortex was less active during the exposure phase for the patterned condition relative to the random condition. This effect was specific to the appearance construction, in which the narration afforded expectations about the video, and was not observed for the approach construction, in which the video was self-explanatory. In Experiment 2, we found similar occipital deactivation during the test phase for congruent vs. incongruent trials with correct responses. The departure construction, like the appearance construction in Experiment 1, allowed participants to generate expectations about the event that was likely to unfold in the video. Critically, the visual input for patterned and random runs (Experiment 1), and for congruent and incongruent trials (Experiment 2) was identical, demonstrating that the narration impacted visual processing.

Occipital deactivation has been implicated in two processes that may be related to the comprehension of constructional meaning: visual categorization and predictive processing. In the categorization literature, the effect has been observed in fMRI studies of prototype extraction from dot patterns (cf. Posner & Keele, 1968; Reber, Gitelman, Parrish, & Mesulam, 2003). There

is significant deactivation in visual cortex for category members (old exemplars or distortions from the prototype) vs. non-members (see also Reber, Stark, & Squire, 1998). Similar findings have also been reported for more abstract categories, such as animals (Chao, Weisberg, & Martin, 2002). Although an appealing account of our findings, the fact that occipital deactivation did not occur for the approach construction — which could have involved categorization processes — suggests that this is not the full explanation.

In the predictive coding literature, occipital deactivation in fMRI studies and single-unit recordings has been linked to the predictability of visual stimuli (Kok, Jehee, & de Lange, 2012; Meyer & Olson, 2011; Summerfield & Egner, 2009). In one compelling example (Kok, Jehee, & de Lange, 2012), participants viewed sequences of two line gratings and were asked to judge either whether the second grating was oriented clockwise or counterclockwise with respect to the first, or whether the second grating was of higher or lower contrast than the first. Each trial began with an auditory tone that predicted with 75% validity the overall orientation of the gratings. When the gratings matched this cue, activation was reduced in occipital cortex, relative to when they did not match. Furthermore, multivoxel pattern analysis was used to demonstrate that such relative deactivation was positively correlated with behavioral performance: Expected visual stimuli received less low-level visual processing as measured by overall activation, but this deactivation reflected a sharpening of the representation that influenced behavior. More generally, this kind of repetition attenuation or suppression in occipital cortex has been linked to perceptual priming (Grill-Spector, Henson, & Martin, 2006; Schacter, Wig, & Stevens, 2007; Turk-Browne, Scholl, & Chun, 2008).

The present studies extend our current understanding of predictive effects in visual cortex. We found that occipital deactivation can occur even when predictions are quite abstract, rather than about a specific low-level stimulus. For the appearance construction, participants could only predict that one entity would appear in or on another entity. For the departure construction, participants could only predict that a named entity would depart. The direction and manner of appearance/departure were novel for each stimulus. That is, the retinal images themselves could not be predicted, only higher-level identities, relations, and motion patterns. The neural sources of these predictions, and why they are realized in low-level visual areas — V1/V2 based on a probabilistic atlas (Wang, Mruczek, Arcaro, & Kastner, 2012) — awaits further study.

Conclusions

There is a growing body of evidence that prediction plays a key role in language production and comprehension (for a review, see Pickering & Garrod, 2013). The present studies provide new evidence that the abstract meanings of grammatical constructions are learned rapidly from minimal exposure, and are quickly used to predict upcoming visual stimuli. Although there have been numerous neuroimaging studies on various aspects of language processing, this work represents the first neuroimaging study of how meaningful novel phrasal constructions are processed in the context of visual scenes.

Experiments 1 and 2 used different stimuli and different designs, yet they converged to indicate an important role for prediction. Attenuated activity in occipital cortex is known to occur when a visual stimulus is predictable, and in the context of grammatical constructions, we found such deactivation when and only when comprehension of a construction enabled prediction of upcoming visual events. In addition, we found that behavioral evidence of learning correlated inversely with activity in the ventral striatum, an area long associated with prediction

error. We suggest that both effects stem from learners rapidly creating and evaluating predictions about the interpretation of instances of grammatical constructions.

Our focus in these studies (especially the hypothesis-driven Experiment 2) was on establishing a role for prediction in the acquisition and use of novel constructions. However, other processes and brain regions also surely play a role. For example, how more specialized language areas and networks (see Fedorenko & Thompson-Schill, 2014) contribute to the processing of grammatical constructions remains an open question. Indeed, although because of careful experimental design and behavioral measures of learning we believe that our occipital and ventral striatal effects are related to predictive processes, these regions have diverse functions and support other processes as well. Until such processes are identified as relevant, however, prediction provides a parsimonious account of our findings.

Acknowledgements

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20,* 952–964.

Alishahi, A., & Stevenson, S. (2007). A computational usage-based model for learning general properties of semantic roles. In *Proceedings of the 2nd European Cognitive Science Conference* (Vol. 425, p. 430).

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, *116*(1), 143–148.

Arbib, M., & Lee, J. (2008). Describing visual scenes: towards a neurolingistics based on construction grammar. *Brain Research*, *1226*, 146–162.

Arnold, J., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language*, *76*, 28–55.

Bates, E., & MacWhinney, B. (1987). in *Mechanisms of Language Acquisition* (pp. 157-193). B. MacWhinney (Ed.). Mahwah, NJ: Erlbaum.

Bloom, L. (1975). *Language development*. University of Chicago Press.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition, 113*, 262–280.

Boyd, J. K., Gottschalk, E. A., & Goldberg, A. E. (2009). Linking rule acquisition in novel phrasal constructions. *Language Learning, 59*, 64–89.

Braine, M. D., Brody, R. E., Brooks, P. J., Sudhalter, V., Ross, J. A., Catalano, L., & Fisch, S. M. (1990). Exploring language acquisition in children with a miniature artificial language: Effects of item and pattern frequency, arbitrary subclasses, and correction. *Journal of Memory and Language*, *29*(5), 591-610.

Bromberg-Martin, E., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron, 63,* 119–126.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500-508.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex, 12*, 545–551.

Chow, H. M., Kaup, B., Raabe, M., & Greenlee, M. W. (2008). Evidence of fronto-temporal interactions for strategic inference processes during language comprehension. *NeuroImage, 40*, 940–954.

Christiansen, M. H., & Chater, N. (1994). Generalization and connectionist language learning. *Mind & Language, 9,* 273–287.

Clark, E. V. (1995). *The lexicon in acquisition* (Vol. 65). Cambridge University Press.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114,* 356–371.

Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

Dominey, P. F. (2002). Conceptual grounding in simulation studies of language acquisition. *Evolution of Communication*, *4*(1), 57-85.

Dominey, P. F. (2003). Learning grammatical constructions from narrated video events for human–robot interaction. In *Proceedings IEEE Humanoid Robotics Conference, Karlsruhe, Germany*.

Dominey, P., & Boucher, J. (2005). Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence*, *167*, 31–61.

Dominey, P. F., Hoen, M., & Inui, T. (2006). A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*, *18*, 2088–2107.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences, 8*, 301–306.

Fedorenko, E., Nieto-Castanon, A., & Kanwisher, N. (2012). Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*, *50*, 499–513.

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences, 18*, 120–126.

Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. *Cerebral Cortex*, *10*(7), 698–705.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684-691.

Gibson, E., Bergen, L. & Piantadosi, S. (2013). Rational integration of noisy evidence and prior expectations in sentence interpretation. *PNAS* 110, 8051-8056.

Glenberg, A. M., & Gallese, V. (2012). Action-based language: a theory of language acquisition, comprehension, and production. *Cortex*, *48*, 905–922.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2005). The role of prediction in construction-learning. *Journal of Child Language, 32,* 407-426.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15,* 289–316.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10,* 14–23.

Grodzinsky, Y., & Friederici, A. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology, 16,* 240–246.

Hagoort 2003; How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, *20*, S18-S29.

Haspelmath, M. (2008). Parametric versus functional explanations of syntactic universals. *The limits of syntactic variation*, 75-107.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In McDaniel, D., McKee, C., & Smith Cairns, H. (Eds.). *Methods for Assessing Children's Syntax. Language, Speech, and Communication*. (pp. 105–124). Cambridge, MA, US: The MIT Press,

Hoen, M., Pachot-Clouard, M., Segebarth, C., & Dominey, P. F. (2006). When Broca experiences the janus syndrome: An ER-fMRI study comparing sentence comprehension

and cognitive sequence processing. *Cortex*, *42*, 605–623. doi:10.1016/S0010-9452(08)70398-8

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language 127*, 46–54.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General, 133,* 450–467.

Kemmerer, D. (2006). Action verbs, argument structure constructions, and the mirror neuron system. In M. Arbib (Eds.), *Action to Language via the Mirror Neuron System* (pp. 347–373). Cambridge University Press.

Kersten, A. W., & Earles, J. L. (2001). Less really is more for adults learning a miniature artificial language. *Journal of Memory and Language*, *44*(2), 250-273.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7,* 376–395.

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75,* 265–270.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161–163.

Lieberman, M. D., Chang, G. Y., Chiao, J., Bookheimer, S. Y., & Knowlton, B. J. (2004). An event-related fMRI study of artificial grammar learning in a balanced chunk strength design. *Journal of Cognitive Neuroscience, 16,* 427–438.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244,* 522–523.

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences USA, 108,* 19401–19406.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science, 2,* 138–153.

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Science, 12,* 265–272.

Noordzij, M. L., Neggers, S. F., Ramsey, N. F., & Postma, A. (2008). Neural correlates of locative prepositions. *Neuropsychologia*, *46*, 1576–1580.

O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304,* 452–454.

Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience, 24,* 8436–8440.

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience, 5,* 97–98.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–225.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11,* 105–110.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36,* 329–347.

Posner, M. I., Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363.

Pulvermüller, F., Cappelle, B., & Shtyrov, Y. (2013). Brain basis of meaning, words, constructions and grammar. In T. Hoffmann & G. Trousdale (Eds.), *Oxford Handbook of Construction Grammar* (Vol. 112, pp. 167–179). Oxford: Oxford University Press.

Reber, P. J., Gitelman, D. R., Parrish, T. B., & Mesulam, M. M. (2003). Dissociating explicit and implicit category knowledge with fMRI. *Journal of Cognitive Neuroscience, 15,* 574–583.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences USA, 95,* 747–750.

Redgrave, P, & Gurney, K. (2006). The short-latency dopamine signal: A role in discovering novel actions? *Nature Reviews Neuroscience, 7,* 967–975.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: current research and theory*.

Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinions in Neurobiology, 17,* 171–176.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Skosnik, P. D., Mirza, F., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2002). Neural correlates of artificial grammar learning. *NeuroImage, 17,* 1306–1314.

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687-E4696.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13,* 403–409.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253.

Traxler, M. J., Foss, D. J., Seely, R. E., Kaup, B., & Morris, R. (2000). Processing effects in sentence comprehension: Effects of association and integration. *Journal of Psycholinguistic Research, 29*, 581–595.

Turk-Browne, N. B., Scholl, B. J., & Chun, M. M. (2008). Babies and brains: Habituation in infant cognition and functional neuroimaging. *Frontiers in Human Neuroscience, 2,* 16.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21,* 1934–1945.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

Ungless, M. A. (2004). Dopamine: The salient issue. *Trends in Neuroscience, 27,* 702–706.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2012). Visual topographic probability maps (VTPM) in standard MNI [and surface] space. *Society for Neuroscience Program* 573.06.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology, 56,* 165–209.

Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5year olds. *Journal of Memory and Language*, *66*(3), 458-478.

Xue, G., Chen, C., Jin, Z., & Dong, Q. (2006). Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. *Neuroimage*, *31*(3), 1315-1326.