Title
Dynamic Action Facilitates Learning of Non-Adjacent Dependencies in Visual Sequences

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Lu, Helen Shiyang
Mintz, Toby

Publication Date
2021

Peer reviewed

Dynamic Action Facilitates Learning of Non-Adjacent Dependencies in Visual Sequences

Helen Shiyang Lu (LUSHIYAN@Usc.Edu)
Department of Psychology, SGM 501, 3620 S.McClintock Avenue
Los Angeles, CA 90089-1061 USA

Toben H. Mintz (TMINTZ@Usc.Edu)
Department of Psychology, SGM 501, 3620 S.McClintock Avenue
Los Angeles, CA 90089-1061 USA
Department of Linguistics, GFS 301, 3601 Watt Way
Los Angeles, CA 90089-1693 USA

Abstract

Many events that humans and other organisms experience contain regularities in which certain elements within an event predict certain others. While some of these regularities involve tracking the co-occurrences between temporarily adjacent stimuli, others involve tracking the co-occurrences between temporarily distant stimuli (i.e., non-adjacent dependencies, NADs). Prior research shows robust learning of adjacent dependencies in humans and other species, whereas learning NADs is more difficult, and often requires support from properties of the stimulus to help learners notice the NADs. Here we report on four experiments that examined NAD learning from various types of visual stimuli. The results suggest that continuous movements aid the acquisition of NADs. We also found that human motion leads to more robust NAD learning compared to object motions, perhaps because of a richer representation. This richer representation could result in better memory and recall, and provide a stronger signal for NAD learning.

Keywords: non-adjacent dependency; human action; visual sequence; statistical learning

Introduction

Many events we experience involve temporally ordered sequences. These include visual events, such as watching agents engaging in actions, as well as auditory events, such as hearing a sequence of words in a spoken sentence. In many cases, these sequential events contain regularities in which certain elements within predict certain others. For example, in the English present progressive, the copula, *is*, is followed by a verb with the inflection *-ing* (e.g., ... ***is** bak-**ing*** ...). Through experience, individuals learn about aspects of these regularities, and, once noticed, can use them to generate new knowledge, either explicit, such as the understanding of an artifact's function, or implicit, such as the knowledge of the grammatical rules in one's native language(s).

There is now a large body of research investigating the kinds of regularities that learners can detect and learn from sequential stimuli. Many studies have looked at regularities involving adjacent items. The findings show that co-occurrence patterns of adjacent items are readily detected by human adults (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), infants (Saffran, Aslin, & Newport, 1996; Aslin, Saffran, & Newport, 1998), and even by other species (Hauser, Newport, & Aslin, 2001; Toro & Trobalón, 2005). The ability to detect adjacent regularities has been shown for a range of stimuli, including speech (Saffran et al., 1996; Pelucchi, Hay, & Saffran, 2009), tones (Saffran, Johnson, Aslin, & Newport, 1999), non-musical noises (Gebhart, Newport, & Aslin, 2009), shapes (Fiser & Aslin, 2002), and human actions (Baldwin, Andersson, Saffran, & Meyer, 2008). Researchers have also explored regularities involving non-adjacent items, or non-adjacent dependencies (NADs). The findings from studies with adult humans suggest that learning of non-adjacent patterns is less robust, and shows a greater dependency on various properties of the stimulus. For example, Peña, Bonatti, Nespor, and Mehler (2002) first tested whether adults could learn NADs from sequences of nonsense syllables in an artificial language, and found that adults were capable of learning NADs from syllable sequences only when brief 25ms pauses were added between syllable triplets that contained the NAD. Wang, Zevin, and Mintz (2019) expanded this finding and demonstrated that learning NADs from syllable sequences was possible without pauses, if the sequences were presented at a natural speech rate. With respect to non-linguistic auditory stimuli, adults could learn NADs from musical tones and noises when the items constructing the dependency (i.e., the *a* and *b*) were perceptually similar to each other and when the middle item X differed perceptually from them (Creel, Newport, & Aslin, 2004; Gebhart et al., 2009). Similar results were found with learning NADs from visual stimuli such as simple shapes (Turk-Browne, Jungé, & Scholl, 2005). Participants failed to acquire the NADs when the perceptual similarity cues were removed from the sequence, even when brief pauses were added (Gebhart et al., 2009).

It is not surprising that pattern detection and learning is different for adjacent and non-adjacent patterns. Linking adjacent items requires minimal memory resources (although resources are required to store the link), and the relationship itself is very basic—items are either adjacent or they are not. In contrast, detecting non-adjacent relationships requires holding one item in working memory as more items are processed, then linking that item to a subsequent (non-adjacent) item. Beyond this increase in resource demands, the computational problem increases because there are multiple non-adjacent relationships that the learner could consider: While adjacency is limited to one position, non-adjacency is bounded only by the length of the sequence. Moreover, there is evidence that learners learn adjacent patterns even as they are learning non-adjacent ones (Romberg & Saffran, 2013; Wang & Mintz, 2018). Thus, given that the non-adjacent relationship itself is less constrained than adjacency, and given the correlated increase in resource demands, learning non-adjacent dependencies would logically benefit from constraints on the domain over which non-adjacent patterns are processed. The pauses in Peña et al.'s (2002) study plausibly contribute one kind of constraint. Wang et al. (2019) proposed that rhythmic segmentation set up by repetition patterns (when there are no pauses) could be another. Perceptual similarity could serve a similar function by packaging the input into perceptually similar units, over which pattern detection mechanisms could operate. In general, anything that could constrain or bias a learner to segment or group a sequence into smaller sub-sequences could facilitate NAD learning, if the sub-sequences contain the NAD (Wang, Zevin, & Mintz, 2017).

In the visual domain, dynamic movement has been proposed as another property that could constrain and facilitate NAD learning. Li and Mintz (2015) showed that when an object dynamically transformed from a flat, blanket-like shape into different forms, learners could detect NADs involving those transformation movements but did not learn NADs involving pictures of the static endpoints of those transformations.[1] Interestingly, if the dynamic events or static images involved human avatars, then learners were able to detect NADs in both dynamic and static images. Li and Mintz hypothesized that movement provided a way of connecting the items in a sequence into a cohesive package, and that this packaging served to constrain and facilitate the detection of the NADs contained within those elements. They further speculated that in the case of static human images, learners filled in the motion, either because of familiarity with seeing human forms in motion, or because they mapped the avatar positions to their own body, and simulated the motor movements to transition from one pose to the next. Either way, they speculated that dynamic motion (real or inferred) bound the events together and and thus facilitated NAD detection, much like the other signal-driven constraints discussed earlier.

A limitation with Li and Mintz's (2015) findings is that their design did not allow them to rule out the possibility that learners were learning simpler positional regularities, and not NADs. In their experiments, participants saw random sequences of nine NAD triplets during training (e.g., $a_1X_1b_1$, $a_1X_2b_1$, $a_1X_3b_1$, $a_2X_1b_2$, $a_2X_2b_2$, etc.). Then, they were tested on triplets that adhered to the NAD patterns (i.e., NAD triplets) and triplets that deviated from the NAD patterns (i.e., part triplets). In the test materials, NAD triplets had items at the beginning and end positions in their trained positions (e.g., a_1Xb_1 where $a_1\_\_b_1$ constituted a trained NAD, but where the X item was an a or b item in training which hadn't occurred in the medial position), whereas part triplets contained an edge position with an item in an unattested position (e.g., a middle item during training occurred in an edge position in testing). On each test trial, subjects had to pick which of two test items, NAD triplet or part triplet, was most like the training triplets. When stimuli were human avatars (static, or in motion) or objects displaying dynamic movement, subjects chose the NAD triplets. Endorsement of the items with the NADs could have been driven by the recognition of the NADs, but could also have been driven by the fact that the first and last items were in familiar positions. Hence, those findings do not conclusively show that dynamic motion facilitates NAD detection in visual sequences. In fact, Li and Mintz (2015) based their study on a prior study by Endress and Wood (2011) that explored the conditions under which subjects learned adjacent dependencies and positional regularities in human action sequences, so the design was not meant to provide a strong test of NAD learning. Frost and Monaghan (2016) also identified problems with this design for testing NAD learning. They argued that using an item in its unattested position might hinder NAD learning and block generalization to triplets in the test phase. When they modified the design by using novel items in the medial position of test triplets, adults were capable of learning NADs in unsegmented speech, in contrast to the findings of Peña et al. (2002). Thus, it is still an open question whether dynamic motion facilitates the detection of NADs in temporally sequenced visual stimuli. This is what the present study addresses.

In the four experiments here, the test items were either NAD triplets, or *positional triplets*. These were sequences where the first and last items appeared in the same position within the sequence as they did in the triplets in the training materials, but where the non-adjacent relationship was violated. For example, if $a_1\_\_b_1$ and $a_2\_\_b_2$ were trained NADs, then a_1Xb_2 and a_2Xb_1 would be positional triplets. This difference in test items compared to Li and Mintz (2015) allows one to test for subjects' sensitivity to NADs, above and beyond their sensitivity to absolute position.

Experiment 1: Human Actions

Experiment 1 investigated adults' capacity for learning NADs from human actions. Li and Mintz (2015) found that subjects endorsed sequences of human actions that contained attested NADs over ones that did not, but their data is consistent with the possibility that subjects learned only the positions of actions within triplet sequences, and not the relations between them. In this experiment, all the items in all test sequences were in the same positions within the test triplet as they were in the training materials; they differed only on whether they contained a trained NAD or not.

Methods

Participants Forty-five participants were recruited online through Prolific, a crowd-sourcing platform for running

[1]The shapes and transformations were similar to those depicted in Figure 3.

experiments. Participants received \$3.75 for their time in the experiment. Ten participants were excluded for the analysis due to technical issues during the experiment (2) or not passing the attention check (8). The final sample included 35 participants ($M_{age} = 25.37$, $SD_{age} = 6.00$, *range*: 18-41 years).

Materials Fifteen video clips of human actions similar to those in Li and Mintz (2015) and Endress and Wood (2011) were used to create training and testing materials. In each video clip, an animated male human avatar performed one action (e.g., kicking a leg). Each action clip lasted 625ms, and it started and ended with the avatar in a neutral upright position with arms at the sides and head facing forward. This ensured that the action sequence flowed naturally from clip to clip. The midpoints of each action clip depicted the maximum extent of movements (Figure 1). The videos of human actions were created using the animation software Poser (Bondware, Inc., 2015).

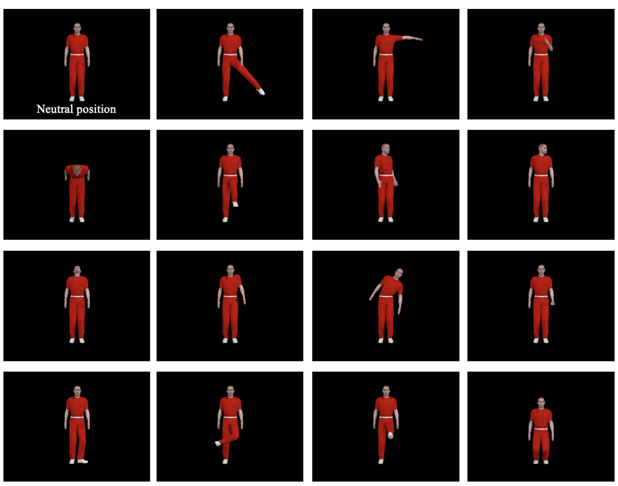

Figure 1: Frames excerpted from the action clips (depicting the maximum extent of movement) used in Experiment 1. These were also the still images of postures used in Experiment 4.

Procedure Participants were first exposed to video sequences of human actions with NADs embedded (i.e., training phase). Six action clips were used to create three NAD frames (i.e., $a_1\_\_b_1, a_2\_\_b_2$, and $a_3\_\_b_3$). For each NAD frame, one of three corresponding intervening actions could be inserted to create a triplet (i.e., X_{1-3} for one frame, and X_{4-6} for a different frame, and X_{7-9} for the third frame). Therefore, participants saw 9 unique triplets during training (i.e., $a_1X_1b_1$, $a_1X_2b_1$, $a_1X_3b_1$, $a_2X_4b_2$, $a_2X_5b_2$, $a_2X_6b_2$, $a_3X_7b_3$, $a_3X_8b_3$, and $a_3X_9b_3$). Furthermore, the assignment of actual actions to letters was random for participants (i.e., a_1 might be raising a leg for one participant and turning head for another). Following Li and Mintz's (2015) design, each triplet was presented for a total of 1875ms, with 125ms pauses between triplets. In the training phase, participants saw a pseudo-random sequence of 20 repetitions of each triplet, resulting in a total of 180 triplets and a total exposure time of 6 minutes.

Then, participants were presented with 36 novel test triplets one at a time. They were asked to indicate for each triplet whether they thought they had seen it before, using a 5-point scale with 1 being definitely had not seen the triplet and 5 being definitely had seen the triplet. Half of the novel triplets followed the NAD patterns (NAD triplets), and the other half violated the patterns (positional triplets). The NAD triplets were created by combining an intervening item of one trained NAD frame with a different NAD frame (e.g., combining the X_5 in $a_2X_5b_2$ and the NAD frame $a_1\_\_b_1$ to create $a_1X_5b_1$). Note that these test sequences were novel since the adjacent transitional probabilities between actions—based on the training sequences—were zero, even while the non-adjacent dependencies were maintained. The positional triplets were created by swapping the last action of one trained NAD triplet with a last action from a different trained NAD triplet, with a middle item that did not occur with either edge item during training (e.g., $a_1X_5b_3$). These test sequences were also novel because the adjacent transitional probabilities were zero; but in this case, the NAD was not maintained. Note that $a_1X_5b_2$ could not be a positional triplet because X_5 had occurred with b_2 in training. Thus, the two types of test items were equated on all surface dimensions with respect to the training items, except the NAD triplets maintained the NAD relationship but the positional triplets did not.

One possible problem with this design was that participants might form a category-like feature for each NAD frame in training and resist generalizing the learned frame with a medial item that went with a different NAD frame in training. This possibility was addressed in Wang et al.'s 2019 study, where the authors compared participants' NAD learning from speech using this design with their learning using a more distributed design (where the nine medial items could go with any of the three NAD frames). Both groups of participants were able to learn the NADs, and showed no difference in their learning outcomes. We therefore did not think that this property would inhibit generalization in our experiments.

Since the experiment was conducted online, twelve catch trials were added into the test phase to ensure participants were paying attention. The catch trials asked participants to rate the familiarity of novel triplets involving repetition of actions (e.g., $a_1b_2b_2$ or $a_1a_1b_2$), which should be obvious to participants that these patterns were novel if they paid attention. Participants who gave a rating of 3 or higher for more than four catch trials were excluded for further analysis. The 48 novel triplets were pseudo-randomly ordered such that no more than 2 triplets of the same type could occur in a row.

If successful NAD learning occurred, we expected participants to rate the NAD triplets higher than the positional triplets. In addition, as the experiment progressed, the difference between participants' ratings for the NAD and positional triplets might become less noticeable. Note that this could be considered a more conservative test of NAD learning than the 2AFC assessment used by Li and Mintz (2015). The difference between NAD and positional triplets is minimal, as the individual items in both test triplet types are in the familiar positions with respect to the training triplets. In 2AFC, subjects compare two items, and the comparison could highlight the small difference between them (the NAD), making the NAD triplet just a little more familiar. In contrast, rating individual test items requires subjects to evaluate each item on its own. Since both test triplet types are very similar to the training triplets, even a small but reliable difference in ratings between NAD and positional triplets is strong evidence for the learning of NADs.

Results and Discussion

An ordinal logistic regression model, with the *ordinal* package in R (Christensen, 2019), was used to compare participants' ratings for the NAD and positional triplets. The model incorporated fixed effects of grammaticality (NAD versus positional), test trial number, and an interaction between the two. In addition, the model allowed for a random intercept for subjects and a by-subject random slope for grammaticality and trial number. Participants' average rating was 3.43 ($SD = 1.24$) for NAD triplets and 3.17 ($SD = 1.30$) for positional triplets. The model indicated that NAD triplets were rated significantly higher than positional triplets ($\beta = .81$, $z = 3.56$, $p < .001$). See Figure 2 for a depiction of participants' ratings of NAD triplets and positional triplets over the period of the test phase. No significant effect was found for trial number ($p = .994$). However, there was a significant interaction between grammaticality and trial number ($\beta = -.02$, $z = -2.29$, $p = .022$). Overall, participants had a smaller difference in their ratings between the NAD and positional triplets as the experiment progressed.

These results show that human adults can learn NADs in temporally sequenced visual events depicting human actions. This findings thus go beyond the findings in Endress and Wood (2011) and Li and Mintz (2015), which showed that subjects can learn positional information in action sequences. Furthermore, participants succeeded even with a more rigorous testing method. Subjects rated each item with respect to its familiarity with the training material, in contrast to some prior methods that used a 2AFC design in which the differences between test items that conform to or violate the trained NADs would be highlighted.

Experiment 2: Object Transformations

Li and Mintz (2015) also investigated NAD learning in sequences of object transformation events. They demonstrated that human adults could learn the positions of items in these sequences. In Experiment 2 we ask whether subjects can learn NADs in such sequences. The question is important because it addresses the generality of the mechanisms involved in visual NAD learning in humans. One could imagine that highly familiar and ecologically important human forms engage learning mechanisms that more generic stimuli do not. Here we use similar object transformations to those used by Li and Mintz, and ask whether human adults can learn NADs in sequences of object transformations.

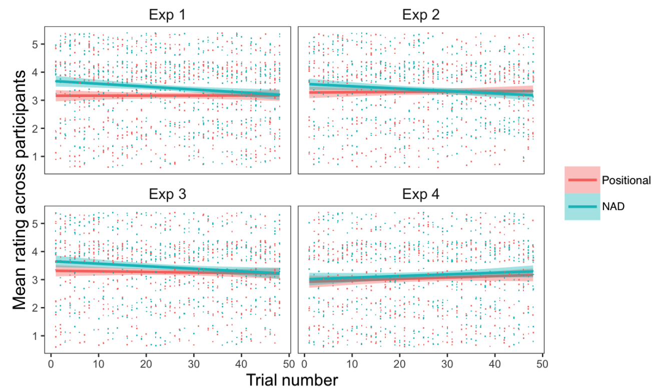

Figure 2: Results of Experiments 1-4. Each graph depicts participants' ratings for positional triplets and NAD triplets over the time course of the experiment. The shades surrounding the lines represent the 95% confidence interval, and each dot represents a rating from one participant for a certain test triplet.

Methods

The methods were the same as in Experiment 1, except that the materials used were video clips of 15 transformations of a blanket-shaped object instead of human actions (Figure 3). These videos were created using Blender (BlenderOnlineCommunity, 2018).

Participants Fifty-eight participants were recruited online through Prolific[2]. Participants received \$3.75 for their time in the experiment. Twenty-eight participants were excluded for the analysis due to technical issues during the experiment (2) or not passing the attention check (26). The final sample included 30 participants (M_{age} = 29.47, SD_{age} = 8.82, *range*: 18-57 years).

Results and Discussion

The same model as the one in Experiment 1 was fitted for Experiment 2. Participants' average rating was 3.37 ($SD = 1.18$) for NAD triplets and 3.30 ($SD = 1.25$) for positional triplets. No significant effect was found for grammaticality ($p = .136$), trial number ($p = .404$), or the interaction between the two ($p = .239$), suggesting that participants failed to distinguish the NAD and positional triplets.

These results differ from those in Li and Mintz (2015), who pitted NAD triplets against part triplets, and found that

[2]All participants only took part in one of the four experiments reported here.

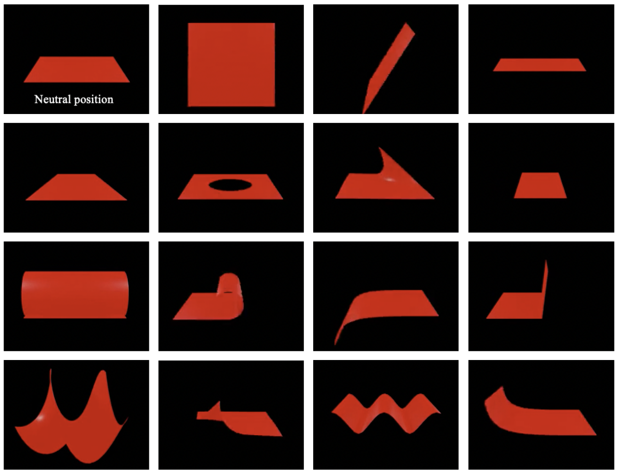

Figure 3: Frames excerpted from the object transformation clips. The image in the upper left corner, labeled "Neutral position," depicts the start and end point of each transformation. Each other image depicts the maximum extent of one of the transformations used in Experiments 2 and 3.

subjects were more likely to choose NAD triplets as the sequences they saw in the training phase. It is possible, as we indicated in the introduction, that subjects in their experiment based their choices on the positional information, and not on the NADs. In our materials, both NAD and positional triplets conformed to the positional patterns in the training triplets. Our different results could also be due to the testing measure we use, as discussed in the Procedure section of Experiment 1. With the 2AFC method Li and Mintz used, subjects always evaluate a 'grammatical' item in comparison to an 'ungrammatical' one, which could draw attention to the critical differences. Here, subjects rate each item individually, and subjects may not always be applying the same standard of comparison in each evaluation. In any case, it is noteworthy that, using the same designs and methods we failed to find evidence of NAD learning in sequences of object transformations, whereas we did in sequences of human actions, in Experiment 1. This contrast suggests that there may indeed be an advantage for human forms in human subjects' visual sequence learning. To compare the learning outcomes across Experiments 1 and 2, we fitted an ordinal regression model with grammaticality, experiment (Experimant 1 vs. 2), and their interaction as the fixed effects, allowing for a by-subject random slope of grammaticality. The lack of an interaction between grammaticality and experiment ($p = .140$) suggests that there might be some weak learning in Experiment 2. Therefore, in Experiment 3, we tested whether increasing the exposure to the training material would lead to detectable NAD learning in sequences of object transformations.

Experiment 3: Object Transformations with More Exposure

In Experiment 1, we found that subjects learned NADs in sequences of human actions, but we failed to find evidence of this capacity when the stimuli were sequences of object transformations. One explanation of this difference is that certain sequence learning mechanisms are only engaged in the domain of human actions, but not for other visual motion events. On the other hand, the apparent advantage for human stimuli may result from greater attention, or richer representations in that domain. In that case, increased exposure to sequences of object transformations could yield NAD learning, even for non-human objects.

Methods

The materials were the same as in Experiment 2, except that the training phase included 30 repetitions of each of the 9 NAD triplets, resulting in an exposure time of 9 minutes. In addition, participants were given a short break after 6 minutes of NAD exposure and saw a short one-minute hand-drawing video during the break.

Participants Forty-five participants were recruited online through Prolific. Participants received \$3.75 for their time in the experiment. Fifteen participants were excluded for the analysis due to technical issues during the experiment (2) or not passing the attention check (13). The final sample included 30 participants (M_{age} = 23.41, SD_{age} = 6.31, *range*: 18-47 years).

Results and Discussion

Participants' average rating was 3.43 ($SD = 1.21$) for NAD triplets and 3.26 ($SD = 1.22$) for positional triplets in Experiment 3. NAD triplets were rated significantly higher than positional triplets ($\beta = .59$, $z = 2.26$, $p = .024$), suggesting that participants learned the NADs embedded in the training phase. No significant effect was found for trial number ($p = .468$) or the interaction between grammaticality and trial number ($p = .094$). Thus, adults indeed have the capacity to learn NADs in sequences of object transformations. However, learning appears to be less robust than in the domain of human actions, as learners required more exposure to the non-human stimuli.

Experiment 4: Human Postures

Li and Mintz (2015) found that subjects could learn positional information (and possibly NADs) from static images of human postures, but not with static images of objects. Thus, the dynamic aspect of the stimuli was not critical for human forms, but was for objects. In our final experiment, we tested NAD learning in sequences of human postures.

Methods

The methods were the same as in Experiment 1, except that each action was substituted with the frame depicting the largest extent of action in that clip (Figure 1). In addition, the twelve attention check triplets had an additional 25ms pause in between postures within a triplet. Otherwise, the attention check triplets might be perceived as a sequence containing only two postures (e.g., instead of $a_1a_1b_2$, participants might perceive it as a longer a_1 followed by a shorter b_2).

Participants Sixty-nine participants were recruited through the University of Southern California Psychology Subject Pool[3]. They received course credits for their participation in the experiment. Thirty-nine were excluded for further analysis because they didn't pass the attention check, leaving 30 participants in the final sample (M_{age} = 20.17, SD_{age} = 1.84, *range*: 18-27 years).

Results and Discussion

Participants' average rating was 3.04 ($SD = 1.32$) for positional triplets and 3.16 ($SD = 1.32$) for NAD triplets in Experiment 4. No significant difference was found in participants' ratings based on grammaticality ($p = .405$), trial number ($p = .131$), or the interaction between the two ($p = .910$), suggesting that participants failed to distinguish the NAD and positional triplets for human postures.

General Discussion

These findings help clarify the role of motion in facilitating learning NADs in temporal visual sequences. Prior research suggested that motion facilitated learning positional information within visual sequences. The present findings extend this result, with evidence of facilitation of NAD learning as well. Moreover, subjects were able to detect action NADs, whether performed by human avatars, or by objects.

Why might motion facilitate NAD learning? In dynamic events, in contrast to static displays, observers witness an entity transform over time. The resulting perception is one in which the temporal dimension is highly salient, and links momentary states or representations of the object, establishing relationships between configurations at different time points. Movement, then , facilitates establishing relationships across time. While any temporal sequence objectively, and necessarily, involves the dimension of time, our hypothesis is that processing dynamic events has the effect of priming, or strengthening the temporal dimension, thus making the relationship between temporally non-adjacent events more accessible. How this is realized at the level of cognitive mechanisms is an important question for future research. We speculate that it may be realized as a kind of chunking or integrative process within working memory. It could also be that dynamic motion events may better sustain people's attention, thus leading to successful NAD learning. Future studies can measure participants' level of attention when viewing NADs from various kinds of events and examine the associated learning outcomes.

We also found that participants didn't learn NADs from sequences of static human poses. This contrasts with what Li and Mintz (2015) found for learning positional information. It may be that our testing method was too conservative to detect learning, as discussed in Experiment 1. In any case, our data do not provide evidence that human forms, as opposed to mere objects, confer any advantage to NAD learning of static images. However, the fact that learners detected NADs in human actions with less exposure than was necessary with object transformations suggests that, as Li and Mintz found with positional learning, human stimuli might confer an advantage for sequence learning. More studies with different objects and transformations would need to be done to rule out the possibility that there were lower-level artifacts of the particular stimuli in Experiments 2 and 3 that made them harder to learn from. One possibility is that some of the object motions in Experiments 2 and 3 looked more complex than the kinds of human motions seen in Experiment 1, in that they involved physical deformations that are rarely seen in the real world. Despite the fact that the human form in Experiment 1 had more parts than the flat plane in Experiments 2 and 3, participants may have a harder time forming mental representations of those complex object motions, resulting in the unsuccessful NAD learning.[4] It would be interesting to investigate whether the advantage of human forms still holds when comparing NAD learning from point-light displays versus NAD learning from groups of dots moving systematically.

It's worth noting that Experiments 1-3 had each motion begin from and return to a neutral position. This feature may aid the segmentation of motions, making it easier to form a mental representations of motions and eventually help NAD learning. On the other hand, it may also make each motion more discrete within a triplet, thus looking less like a cohesive package for NAD learning. Future studies could have the human avatar and the object plane continue to the next motion without returning to a neutral position. This also better mimics the kind of motions we see in the real world, especially those performed by humans.

In summary, we found that subjects can learn NADs in temporally sequenced visual stimuli. Like Li and Mintz (2015), we found that dynamic motion contributed to successful NAD learning, and we also found a learning advantage for human forms. Note that in our experiments, there were brief pauses between triplets during the training phase. Thus, there were already aspects of the signal that were like the pauses in Peña et al.'s (2002) study, yet that was insufficient for NAD learning with static images. Dynamic motion appears to bring the events together in a cohesive package and to stimulate processing and computing relationships over time.

Acknowledgments

This research was supported in part by a University of Southern California Department of Psychology Research Grant to Helen Shiyang Lu, and research funds from the University of

[3]This was a different participant recruitment platform than the one used in the previous three experiments. However, we have recently started running this experiment with subjects recruited from Prolific. Data from 24 subjects show the same patterns as reported here.

[4]We thank an anonymous reviewer for raising this possibility.

Southern California Dornsife College of Letter's, Arts, and Sciences to Toben H. Mintz.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407.

BlenderOnlineCommunity. (2018). *Blender - a 3d modelling and rendering package.* Stichting Blender Foundation, Amsterdam. Retrieved from `http://www.blender.org`

Bondware, Inc. (2015). *Poser.* Retrieved from `https://www.posersoftware.com/`

Christensen, R. H. B. (2019). *ordinal—regression models for ordinal data.* (R package version 2019.12-10. https://CRAN.R-project.org/package=ordinal)

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119–1130.

Endress, A. D., & Wood, J. N. (2011). From movements to actions: Two mechanisms for learning action sequences. *Cognitive Psychology*, *63*(3), 141–171.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458–467.

Frost, R. L., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, *16*(3), 486–490.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition*, *78*(3), B53–B64.

Li, J., & Mintz, T. H. (2015). Constraints on learning non-adjacent dependencies (NADs) of visual stimuli. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 1350–1355). Austin, TX: Cognitive Science Society.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*(5593), 604–607.

Romberg, A. R., & Saffran, J. R. (2013, September). All together now: concurrent learning of multiple structures in an artificial language. *Cognitive Science*, *37*(7), 1290–1320.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105.

Toro, J. M., & Trobalón, J. B. (2005, July). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*(5), 867–875.

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564.

Wang, F. H., & Mintz, T. H. (2018, April). Learning non-adjacent dependencies embedded in sentences of an artificial language: When learning breaks down. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 604–614.

Wang, F. H., Zevin, J., & Mintz, T. H. (2019). Successfully learning non-adjacent dependencies in a continuous artificial language stream. *Cognitive Psychology*, *113*, 101223.

Wang, F. H., Zevin, J. D., & Mintz, T. H. (2017, December). Top-down structure influences learning of nonadjacent dependencies in an artificial language. *Journal of Experimental Psychology: General*, *146*(12), 1738–1748.