Statistical learning of action: The role of conditional probability

Meredith Meyer · Dare Baldwin

Abstract Identification of distinct units within a continuous flow of human action is fundamental to action processing. Such segmentation may rest in part on statistical learning. In a series of four experiments, we examined what types of statistics people can use to segment a continuous stream involving many brief, goal-directed action elements. The results of Experiment 1 showed no evidence for sensitivity to conditional probability, whereas Experiment 2 displayed learning based on joint probability. In Experiment 3, we demonstrated that additional exposure to the input failed to engender sensitivity to conditional probability. However, the results of Experiment 4 showed that a subset of adults—namely, those more successful at identifying actions that had been seen more frequently than comparison sequences—were *also* successful at learning conditional-probability statistics. These experiments help to clarify the mechanisms subserving processing of intentional action, and they highlight important differences from, as well as similarities to, prior studies of statistical learning in other domains, including language.

Keywords Action processing · Action segmentation · Statistical learning · Conditional probability

M. Meyer (✉)
Department of Psychology, University of Michigan,
Ann Arbor, MI 48109, USA
e-mail: mermeyer@umich.edu

D. Baldwin
University of Oregon,
Eugene, OR, USA

We draw inferences and make predictions about others' actions on a daily basis, and these tasks are supported in part by our ability to segment continuous action into discrete units. Past research has demonstrated that one source of information people can use for segmentation lies in the statistical regularities present within an action stream (Baldwin, Andersson, Saffran, & Meyer, 2008; Swallow & Zacks, 2008). A question left unanswered by this research, however, is the precise type of statistical information people capitalize on for segmentation purposes. Prior studies of statistical learning in other modalities have clarified that both infants and adults are sensitive to conditional-probability statistics, a higher-order statistic encoding the likelihood with which one element predicts the presence of another (Aslin, Saffran, & Newport, 1998; Aslin et al. 2001; Fiser & Aslin, 2001, 2002a, 2002b; Graf Estes, Evans, Alibali, & Saffran, 2007). The present experiments addressed whether a similar sensitivity to conditional probability subserves action segmentation. In this study, we also assessed learning of another type of statistic, joint probability. Joint probability is a distinct type of statistic expressing the frequency with which elements co-occur.

Segmentation of action rests fundamentally on the ability to recognize when one action has ended and another has begun. Evidence suggests we usually accomplish this with a high degree of ease and consistency. When asked to note the onsets and offsets of events, people typically report dynamic human action as consisting of units corresponding to initiation or completion of goals, with considerable agreement across individuals regarding where action boundaries are located (Baldwin & Baird, 2001; Hard, Tversky, & Lang, 2006; Zacks, 2004). Further, action segmentation is a seemingly spontaneous, automatic, and relatively effortless process, engaged in as an ongoing

component of perception (Hard, Recchia, & Tversky, 2011; Saylor & Baldwin, 2004; Speer, Swallow, & Zacks, 2003; Zacks & Swallow, 2007; Zacks, Tversky, & Iyer, 2001). The ease with which we recognize action units, however, belies the complexity of the action itself and, by extension, our ability to analyze it. Careful consideration of human action suggests that it is, in fact, an unquestionably complex stimulus. Even mundane, everyday intentional action frequently lacks pauses or other clear markers of when a goal is completed and another initiated; instead, human action unfolds in a fairly fluid manner (Newtson & Enquist, 1976). Further, the visual scene is frequently complicated and busy, with access to a full action unit frequently blocked by the presence of other, occluding objects. The dynamic, complex, and continuous nature of action thus, in principle, poses a considerable challenge for segmentation.

The ability to analyze human action in terms of intentions likely draws heavily from top-down inferences regarding an actor's desires and beliefs, as well as from knowledge of social roles, one's own past experience with the action, object affordances, and other world knowledge (see, e.g., Schank & Abelson, 1977). For example, imagine being in a restaurant and watching a waiter deliver food to a nearby table. Familiarity with the waiter's goal of delivering the correct food to the diners, past experience with actions involved in serving food, and an understanding of the need to balance the plates while serving the diners all aid us in recognizing the onsets and offsets of individual action units. Use of inferences regarding goal states or mental states on the part of the actor are unlikely to account for the entire story behind action segmentation, however; for instance, even infants lacking sophisticated knowledge about goals and intentions are capable of segmenting simple action streams (e.g., Baldwin, Baird, Saylor, & Clark, 2001). A number of researchers have offered ideas on what sources of information might be used by bottom-up mechanisms in the service of segmentation, including such features as movement, color, sound changes, and the

statistical regularities characterizing the unfolding of events in the action stream (e.g., Avrahami & Kareev, 1994; Baird & Baldwin, 2001; Baldwin et al., 2008; Baldwin & Baird, 2001; Hard et al., 2011; Hard et al., 2006; Newtson, Engquist, & Bois, 1977; Tversky, Zacks, & Hard, 2008; Zacks, 2004; Zacks, Kumar, Abrams, & Mehta, 2009).

In the present experiments, we address in detail one of these candidate mechanisms, which is specifically sensitivity to statistical regularities among small motion elements. To illustrate, imagine being asked to identify individual action units performed by a person cooking a meal. Even without extensive prior knowledge of cooking or food preparation, basic statistical learning processes could enable detection of which elements belong to the same action, simply by virtue of being sensitive to the statistical structure characterizing the stream of smaller motion elements (e.g., *grasp knife/chop vegetable* might be two small motion elements that occur with statistical regularity, allowing them to be detected as one unit). Detection of statistical structure could allow for linkage of the smaller parts, promoting segmentation at a higher level of analysis without requiring extensive top-down knowledge of intentions or goals.

In this article, we examine two specific types of statistics that might be tracked—namely, joint probability and conditional probability. Whereas joint probability expresses the likelihood with which multiple events co-occur, conditional probability provides predictive information relating the likelihood with which one event will occur, given the presence of another event. To illustrate how joint and conditional probabilities might differ in the structure of an everyday stream of actions, consider a sequence of kitchen events, including Event A (stirring soup), Event B (tasting the soup), Event C (chopping a vegetable), Event D (rinsing a dish), and Event E (seasoning the soup). Assume that one needs to stir the soup throughout the cooking process, such that Event A (stirring) occurs more frequently than any of the other events.

A	→	B	→	D	→	A	→	C	→	A	→	D	→	E	→	B
stir		taste		rinse		stir		chop		stir		rinse		season		taste

In an event chain such as this one, the joint probabilities $P(A, B)$ and $P(E, B)$ are equal, because AB and EB both occur together one time across the sequence of events. However, the conditional probability of B given E, or $P(B|E)$, is higher than the conditional probability of B given A, or $P(B|A)$, because E is perfectly predictive of B, whereas A is not. That is, all occurrences of E are

followed by B, but A is followed by other elements in addition to B (see also Fiser & Aslin, 2002a, for a similar explanation of the distinction between joint and conditional probabilities).

What implications might this have for action segmentation? It is arguably the case that the actions of seasoning and tasting the soup (Events E and B, respectively) are

more motivated by a single overarching goal of perfecting the taste of the soup, whereas the actions of stirring and tasting the soup (Events A and B) seem less linked by a common goal. A mechanism sensitive to conditional probability would detect seasoning and tasting as being part of a larger unit; on the other hand, a mechanism sensitive only to joint probability would not be able to take into account the predictive information, instead treating *stirring/tasting* and *seasoning/tasting* as equivalent.

The distinction between conditional probability and joint probability has long been a focus of the learning literature more generally, with work in the nonhuman domain demonstrating that animals' associative learning typically reflects conditional-probability learning (i.e., learning of predictive relationships) rather than mere sensitivity to joint probability (i.e., learning of co-occurrence frequencies; see, e.g., Rescorla & Wagner, 1972). More recent work with humans has also examined conditional-probability learning distinct from joint-probability learning, and we review this work in the following section.

Statistical learning in other domains

A largely analogous inquiry into segmentation skills within the language domain has already established that people are sensitive to the conditional probabilities among syllabic units. Present even in infancy, this skill is believed to be one source of information that, among other things, helps infants segment the continuous speech stream into words. In the first demonstration of infants' sensitivity to the statistical structure of syllables, Saffran, Aslin, and Newport (1996a) exposed infants to a series of syllables that occurred in wordlike groups of three ("words"). After listening to these syllables repeated in a continuous speech stream, infants were given the opportunity to listen to sequences whose average conditional probabilities were 1.0 (i.e., a word) or sequences whose average conditional probabilities were less than 1.0. Infants preferred to listen to the latter, suggesting that they deemed these sequences more novel than the words they had segmented during exposure. However, as the authors later noted, infants may also have been responding to another statistical regularity—namely, the co-occurrence frequencies of syllables (Aslin et al., 1998). Specifically, because each word was presented an equal number of times in the exposure corpus, words not only had overall higher conditional probabilities among the syllables that comprised them, but also occurred more frequently than any other trisyllabic sequence. Since the frequency of co-occurrence was higher for syllables comprising words than for other sequences, infants may have been responding to this difference in joint probability rather than computing conditional probabilities.

Aslin et al. (1998) resolved this issue by controlling for the frequency with which test sequences were heard during exposure. They found that after exposure, infants again preferred to listen to the sequences with lower average conditional probabilities, as in the original Saffran et al. (1996a) findings. Crucially, however, the sequences that infants discriminated between had been presented the same number of times during exposure. This study thus unambiguously demonstrated that infants were able to compute conditional probabilities across a continuous speech stream and to discriminate sequences whose conditional probabilities differed.

Related work has demonstrated that adults are also capable of statistical learning. By asking participants to report which of two sequences was more familiar (analogous to the familiarity/preference measures used for the infancy work described above), studies have indicated that adults can use statistical regularities to segment speech (Saffran, Newport, & Aslin, 1996b), musical tones (Saffran, Johnson, Aslin, & Newport, 1999), and visuomotor sequences (Hunt & Aslin, 1998). Although these studies did not directly address whether adults were computing conditional probability, joint probability, or both, other research has suggested that adults are indeed sensitive to conditional probability independent of co-occurrence frequency. Adults compute conditional probabilities both in spatial configurations of shapes (Fiser & Aslin, 2001) and in shape sequences (Fiser & Aslin, 2002a), evidenced by their reporting of sequences characterized by higher conditional probabilities as being more familiar than frequency-balanced comparison sequences. Taken together with the infancy findings of Aslin and colleagues (Aslin et al., 1998; Aslin et al., 2001; Fiser & Aslin, 2002b; see also Graf Estes et al., 2007, for a replication with speech), research has suggested that sensitivity to conditional probability is a robust mechanism that persists across development and can function in multiple modalities.

Statistical learning in action

Prior investigations of statistical learning in other modalities thus seem to suggest that discovery of higher-level action units is likely also enabled by sensitivity to conditional probabilities that relate successive events to one another. Given that conditional-probability learning has been demonstrated across a wide range of modalities (aural and visual), a variety of input types (syllable sequences, shape sequences, and shape configurations), and diverse ages (infant and adult), it seems highly likely that similar learning abilities would exist for action processing. Empirical demonstration of such an ability, however, is nevertheless important, given the implications for current theories of

event processing. For instance, Zacks and colleagues (e.g., Kurby & Zacks, 2008; Zacks, Speer, Swallow, Braver, & Reynolds, 2007) have proposed the *event segmentation theory*, an account of how the human observer perceives and conceptualizes action in terms of events. A crucial component of event segmentation theory rests on the observer's ability to make predictions about upcoming actions. Such prediction generation is considered an implicit, spontaneous, and online process that integrates incoming sensory information with prior knowledge and learning in an attempt to create a stable "event model." Zacks and colleagues (e.g., Zacks et al., 2009; Zacks et al., 2007) have described a variety of cues, both top down and bottom up, that might feed into such a predictive system; importantly for the purposes of this article, statistical information capturing sequential dependencies among events is one such proposed source of information. Specifically, Zacks and colleagues theorized that information about statistical regularities in action would be incorporated as one component of a relatively stable model, which in turn would influence a person's online processing of an event.

In event segmentation theory, prediction is proposed to be uniquely important in signaling the onsets and offsets of individual events. Event units correspond to periods in which predictability is high, the observed action is consistent with the predictions being made by the processing system, and the event model is stable. For example, within the event of cleaning off plates at the kitchen sink, the predictive system is able to generate accurate predictions of further plate cleaning based on such cues as the person's movements and prior knowledge about kitchen clean-up. Event boundaries, in contrast, are experienced when predictability is low; to extend the example above, such boundary moments are likely to occur at the completion of a task (e.g., cleaning off plates in the kitchen) and before the initiation of another task (e.g., wiping the countertop), because these moments correspond to a reduced ability to predict the onset of the second event. Importantly, predictability would thus partly be determined by action that was inconsistent with prior knowledge of statistical regularities; when a stream of action is consistent with previously learned statistical information (e.g., cleaning a plate would typically be followed by cleaning another plate), predictability would be high, whereas seeing action inconsistent with this statistical information (e.g., cleaning a plate would not typically be followed by wiping a countertop) would lead to the perception of a boundary. A system fundamentally dedicated to prediction would likely benefit from conditional-probability information, as this type of statistic is itself reflective of the predictive relationships among elements. It remains an open question, however, what type of statistical information is actually computed by the human observer of actions.

Explorations of conditional-probability learning for actions also bear on issues related to statistical learning more generally. Debate exists regarding the degree to which statistical learning reflects a single, domain-general mechanism. Some researchers have pointed to findings of similar learning abilities across modalities as support for a domain-general approach. For example, Kirkham, Slemmer, and Johnson (2002) suggested that evidence found in prior research for visual statistical learning in infants mirroring auditory learning (e.g., Saffran et al., 1996a) is evidence for a domain-general mechanism. On the other hand, dissociations in learnability depending on modality have also been discovered and have been used to argue for modality-specific mechanisms. For example, Conway and Christiansen (2005) compared learning success for artificial grammars across three different modalities: touch, vision, and audition. While they found learning in all three modalities, learning was superior in the auditory modality; furthermore, participants were more sensitive to items presented near the end of exposure, relative to those near the beginning of exposure, for auditory learning, whereas the opposite was true for tactile learning. The authors argued that this constitutes evidence for separate, modality-specific mechanisms. In a similar vein, other studies have suggested that rule learning of speech is easier than learning in other modalities and that speech is thus privileged, with the learning mechanism itself possibly adapted for speechlike input (Marcus, Fernandes, & Johnson, 2007).

One question that arises from these findings that is particularly germane to this debate is whether mechanism differences per se contribute to the observed dissociations across modalities, or rather whether a single mechanism produces different outputs based on input stimulus differences. This question was, for instance, integral for Gebhart, Newport, and Aslin (2009), who examined learning for both adjacent and nonadjacent dependencies in auditory nonspeech noise learning. The striking similarity between the overall patterns of learning in their study and the patterns seen in past studies of speech learning led these researchers to suggest that statistical learning likely operates fundamentally similarly across a range of materials. These researchers also found, however, that participants required much more exposure to the nonspeech noise elements in order to learn the statistics, relative to the exposures required in past studies of speech learning. Gebhart et al. attributed this difference *not* to the existence of separate mechanisms, though, but rather to the unfamiliarity or reduced encodability of the nonspeech stimuli (see also Saffran, Pollak, Seibel, & Shkolnik, 2007, for a similar argument in rule learning). While our study is not intended to directly address the issue of the domain specificity of statistical learning, discerning fundamental similarities and differences in the processing of other types of stimuli, such as human actions, is an important contribution to the issues outlined above.

Overview of the present study

Past work in the action domain has already provided some initial findings regarding people's ability to detect statistical regularities in human actions. Baldwin et al. (2008) studied adults' segmentation of a novel stream of action, making use of a methodology similar to that already used in speech (e.g., Saffran et al., 1996b), and they found largely similar patterns of learning for actions relative to what had been demonstrated for speech and other modalities. Because the present study employs a very similar methodology, a more in-depth description of their stimuli is in order. Participants watched a sequence of continuous human object-directed action featuring 12 small motion elements (SMEs). SMEs were grouped into four sequences, henceforth referred to as *actions*, with each action consisting of three SMEs (e.g., Action 1, *stack/poke/drink*; Action 2, *blow/touch/rattle*). Actions were randomly ordered in a continuous fashion to construct an exposure corpus. The conditional probability among SMEs within an action was 1.0 (e.g., for Action 1, *stack* was always followed by *poke*, which in turn was always followed by *drink*). However, when adjacent elements crossed action boundaries, creating a *part action*, the average conditional probability decreased (e.g., the conditional probability among *rattle/stack/poke* was on average lower, because *rattle* was not perfectly predictive of *stack*—that is, $P(stack \mid rattle)$ was less than 1.0). During test, participants were shown action/part-action pairs and asked to determine which was more familiar. Participants displayed a systematic tendency to select actions as more familiar than part actions, suggesting that they had segmented the stream based on the statistical regularities inherent in the action stream.

Baldwin et al. (2008) provided a clear demonstration that people can use statistical learning to segment human action. Baldwin and colleagues' design, however, like that of many other statistical learning studies (including Saffran et al., 1996a, as described above), did not control for the frequency of co-occurrence of small motion elements comprising the actions versus the part actions. Rather, sequences with higher conditional probabilities (actions) also occurred more frequently during exposure than the comparison part-action sequences. It is thus ambiguous whether, in making judgments between actions and part actions, adults were responding based on their computation of conditional probabilities among small motion elements, or rather to the higher joint probabilities (co-occurrence frequencies) of the small motion elements comprising the actions.

We addressed this ambiguity in a series of four experiments in order to determine the type of statistic—conditional and/or joint—that people are capable of using for action segmentation. In all experiments reported in this article, participants watched an exposure corpus of continuous action consisting of concatenated three-unit *actions* (e.g., Action 1: *stack/poke/drink*). During a subsequent test phase, three-unit sequences were presented in pairs (actions vs. part actions in Experiments 1, 3, and 4, and actions vs. other actions in Exp. 2). Participants were then asked to report which sequences were more familiar. In Experiment 1, participants were exposed to action sequences constructed in such a way as to control for co-occurrence frequency during test (cf. Aslin et al., 1998). In this way, we were able to investigate participants' ability to segment solely based on conditional probability. In Experiment 2, we probed participants' sensitivity to a different type of statistical regularity—namely, joint probability information available from co-occurrence frequencies of small motion elements. To anticipate, participants did not appear to learn conditional-probability statistics in Experiment 1, but they did show evidence of sensitivity to joint probability. Thus, we designed Experiment 3, which was similar to Experiment 1 in assessing conditional-probability learning, but this time provided participants with more exposure to the statistically structured input. Finally, Experiment 4 assessed both joint- and conditional-probability learning in a within-subjects design, enabling us to address whether individual performance on both types of learning was related.

Experiment 1

Experiment 1 controlled for the frequency of actions and part actions presented during test, enabling an assessment of segmentation based solely on sensitivity to conditional probability. Specifically, it featured an exposure corpus in which half of the actions were presented twice as frequently as the other two, allowing for selection of a subset of part actions (i.e., part actions consisting of elements from the more frequent actions) that occurred with the same frequency as the low-frequency actions. Test trials featured these part-action sequences compared with the low-frequency actions; these pairings thus featured identical joint probabilities (co-occurrence frequency), but the part actions had lower average conditional probabilities than the actions. The selection of actions as more familiar than part actions would thus imply sensitivity to conditional-probability statistics.

Method

Participants A total of 32 students at a large Northwest university (24 female, 8 male) received class credit for participation.

Materials Following previous studies of human action segmentation (e.g., Baldwin et al., 2008), we filmed 12 individual object-directed motions, termed *small motion elements* (SMEs). Each individual SME featured a female actress manipulating a glass bottle (see Table 1 for the full

**Table 1** Small motion elements (SMEs) in Experiments 1, 2, 3 and 4

SME	Description
Empty	bottle is turned over as if to pour into open hand
Clean	flat hand wipes top of bottle
Under	bottom of bottle is examined
Feel	index finger touches side of bottle in an up-and-down motion
Blow	bottle is lifted to mouth and blown into
Look	bottle is lifted to face and interior examined
Drink	bottle is lifted and tipped into mouth as if drinking
Twirl	bottle edge is lifted from table and spun around
Read	finger traces over label and bottle is lifted from table as if to read
Rattle	bottle is lifted close to ear and shaken
Slide	bottle is pushed forward on the table and returned
Poke	index finger is inserted and removed from top of bottle

list). Each SME started and ended with the actress in the same position, enabling concatenation of SMEs in any order to result in the appearance of continuous and physically plausible intentional motion. SMEs were grouped into four *actions*, with each action consisting of three randomly selected SMEs (e.g., in one exposure corpus, the four actions were Action 1, *empty/clean/under*; Action 2, *feel /blow/look*; Action 3, *drink/twirl/read*; and Action 4, *rattle/slide/poke*).

In order to enhance the continuity of the motion stream, transitions between individual SMEs, both within and across actions, were smoothed using the Overlap transition in iMovie (Version 5.0.2). SMEs were also doubled in speed in order to create a corpus length that was manageable for our participants. Doubling the speed in this way yielded an action stream that appeared natural, though at a high rate of speed. We then created an exposure corpus

(Corpus 1) approximately 24 min long that contained 120 (i.e., high-frequency) tokens of two actions, Actions 1 and 2, and 60 (i.e., low-frequency) tokens of the other two, Actions 3 and 4. Actions were randomly ordered, with the exception that no action could follow itself. We also identified certain sequences for use in test called *part actions*, which were sequences of three SMEs that spanned the action boundaries (see Fig. 1 for an example).

It should be noted that while the SMEs themselves had clear underlying intentions and were thus meaningful (e.g., stack, poke, drink), the actions they comprised (e.g., *stack/poke/drink*) were arbitrary and without obvious intentional content at the higher-order level of the triad. Put another way, actions contained no greater degree of intentional content than did part actions, meaning that rich top-down knowledge regarding intentional content was not available to aid recognition of action segments. However, in order to
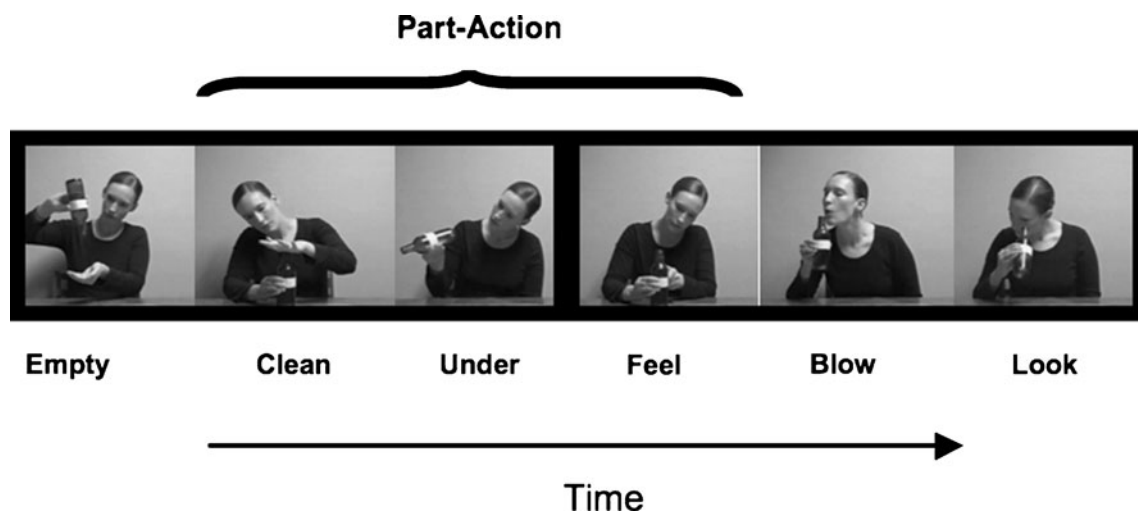
Fig. 1 Still frames excerpted from a sample portion of the continuous action stream from Exposure Corpus 1. The actions displayed (in black frames) are *empty/clean/under* and *feel/blow/look*; the bracketed part action *clean/under/feel* spans an action boundary

avoid the concern that certain sequences were, just by chance, a priori more readily segmentable, we created three more exposure corpora using a control implemented in past studies to address the same concern (e.g., Aslin et al., 1998; Baldwin et al., 2008; Saffran et al., 1996a). Namely, actions from Corpus 1 served as part actions in Corpus 2, and vice versa. We further counterbalanced which actions occurred with high frequency and which with low frequency during exposure, such that high-frequency actions in Corpus 1 were low-frequency actions in Corpus 3 (and vice versa; i.e., low-frequency actions in Corpus 1 were high-frequency actions in Corpus 3), and high-frequency actions in Corpus 2 were low-frequency actions in Corpus 4 (and vice versa; i.e., low-frequency actions in Corpus 2 were high-frequency actions in Corpus 4).[1] The actions and part actions from all corpora, as well as their frequencies, are listed in Table 2.

Pairs of SME triads were selected for discrimination during test. In each pair, one sequence was an action, and the other was a part action (i.e., a sequence that spanned an action boundary). Recall that in previous studies of action segmentation (e.g., Baldwin et al., 2008), part actions featured both lower average conditional probabilities and lower frequencies than their comparator actions, making it impossible to determine the type of statistic that allowed participants to differentiate actions from part actions. In the present study, the frequency difference between high- and low-frequency actions (e.g., Actions 1 and 2 vs. Actions 3 and 4 in Corpus 1) allowed us to control for SME co-occurrence frequency by selecting only actions and part actions that were likely to have equal numbers of occurrences during exposure (see Aslin et al., 1998, for a similar methodology). That is, we selected part actions consisting of elements from the high-frequency actions (e.g., the part-action sequence *clean/under/feel* comprised elements from the high-frequency actions *empty/clean/under* and *feel/blow/look* in Corpus 1); these part actions were likely to occur 60 times during exposure, the same frequency as the low-frequency actions. (By way of example, consider the 120 occurrences of the high-frequency action *empty/clean/under*. Half of the time, this was likely to be followed by the high-frequency action *feel/blow/look*, a quarter of the time by the low-frequency action *drink/twirl/read*, and a quarter of the time by the other low-frequency action, *rattle/slide/poke*. Thus, the predicted frequency of the part-action sequence *clean/under/feel* was 60, equal to the frequencies of the low-frequency actions.) We exhaustively paired the two frequency-balanced part actions with the two low-frequency actions to create four frequency-balanced test trials. We then reversed the order of presentation of each of these four pairings, to produce a total of eight test trials. Test trials were randomly ordered, with the exception that no test trial could directly follow its reverse-order counterpart (e.g., a trial comparing *drink/twirl/read* and *clean/under/feel* could not be followed by a trial comparing these same clips in the reverse order—that is, *clean/under/feel* and *drink/twirl/read*).

Procedure Participants were randomly assigned to one of four exposure corpora and instructed to watch the corpus. In order to ensure their attention to the exposure corpus, participants were told that they would be asked questions about what they saw after the exposure; however, no further information was provided about the nature of the task, and thus any statistical learning that might occur was unsupervised. Immediately after the end of the exposure corpus, we provided participants with two practice forced choice trials to accustom them to the testing format. Practice trials featured a pair of action sequences in succession (separated by a black screen displayed for 1,500 ms). The trials were separated by a 5,000-ms interval in which participants saw written instructions on the screen prompting them to make a response. The test actions were similar in length to the actions and part actions, but the sequences used in the practice trials were entirely different from those seen during exposure and were performed by a different actor. Participants were instructed simply to choose one of the two action clips at the appropriate prompted response time, basing their decision on any standard they wished. Immediately after the practice trials were over and participants' understanding of the testing format was verified, the actual test phase began, and we asked participants to identify which of two clips was more familiar to them based on their previous viewing of the exposure corpus.

Results and discussion

Segmentation based on conditional probability would be demonstrated by greater-than-chance selection of actions as more familiar than part actions. Since the actions and part actions presented during test were frequency balanced, any systematic selection of actions would not be possible based on joint probability information, but instead could only be enabled by conditional probability

---

[1] As a final check that certain sequences were not perceived as more coherent or intentional than others, at the end of the session we asked a subset of the participants from Experiment 4 ($n = 25$) whether they had noticed any sequences as being more coherent than others. The majority of participants (60%) answered in the negative; the remaining participants mentioned certain sequences, but excepting two reports that *clean* cohered well with *drink*, the sequences that participants generated differed from one another. Thus, no sequences were strongly identified as appearing more coherent than others, alleviating concerns regarding the possibility of a priori segmentation of certain sequences.

Table 2 Frequencies and conditional probability information for actions and part actions in Corpora 1 and 3 and Corpora 2 and 4[b]

Actions and Part Actions	Frequencies	Conditional Probability[a]	Frequencies	Conditional Probability[a]
Corpora 1 and 3 Actions	Corpus 1		Corpus 3	
Empty/clean/under	120		60	
Feel/blow/look	120		60	
Drink/twirl/read	60		120	
Rattle/slide/poke	60		120	
Corpora 1 and 3 Part Actions	Corpus 1		Corpus 3	
Clean/under/feel	60	0.5	12	0.2
Blow/look/empty	60	0.5	12	0.2
Read/rattle/slide	12	0.2	60	0.5
Poke/drink/twirl	12	0.2	60	0.5
Corpora 2 and 4 Actions	Corpus 2		Corpus 4	
Clean/under/feel	120		60	
Blow/look/empty	120		60	
Read/rattle/slide	60		120	
Poke/drink/twirl	60		120	
Corpora 2 and 4 Part Actions	Corpus 2		Corpus 4	
Empty/clean/under	60	0.5	12	0.2
Feel/blow/look	60	0.5	12	0.2
Drink/twirl/read	12	0.2	60	0.5
Rattle/slide/poke	12	0.2	60	0.5

[a] Only conditional probabilities < 1.0 are provided within the table. These statistics refer to the likelihood with which an element *within* an action would appear with the element *outside* the action—for instance, for the part action *clean/under/feel* in Corpus 1, in which *clean* and *under* are part of one action and *feel* is part of a separate action, $P(feel \mid under) = .5$. Elements that are part of the same action predict one another perfectly (i.e., feature conditional probabilities of 1.0).

[b] The frequencies of part-action sequences provided here are *predicted* based on the frequencies and ordering of the action sequences; because sequencing in the exposure corpora was accomplished via randomization, actual frequencies of part actions varied slightly from those given here.

computation. Across the eight test trials, however, participants did not discriminate actions from part actions. Mean action selection did not differ from chance levels ($M = 57.81\%$, $SD = 30.40$), $t(31) = 1.45$, $p = .16$ (see
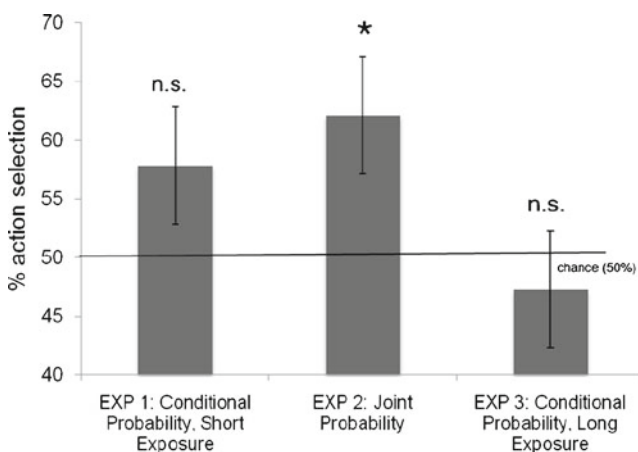
**= different from chance, $p < .05$

Fig. 2 Mean action selections in Experiments 1, 2 and 3. The error bars represent ± 1 standard error. *Different from chance, $p < .05$

Fig. 2). Additionally, only 17 of the 32 participants selected actions more frequently than part actions (i.e., five or more out of eight trials), which was not significant by a binomial test, $p = .86$.[2]

Results from Experiment 1 indicated that participants did not reliably differentiate actions from part actions based on conditional probability. These null results raise the possibility that previous findings of adults' action segmentation were due to participants' sensitivity to joint-probability information. That is, in prior demonstrations of statistical learning of action, participants may have recognized actions as more familiar based simply on the fact that they had occurred more frequently during exposure, rather than computing the conditional probabilities across sequences. Experiment 2 provided a direct test of whether adults are indeed capable of computing joint-probability information in continuous action and using it in the service of segmentation.

---

[2] As expected, one-way ANOVAs demonstrated that performance did not significantly differ across exposure corpora in this experiment or in Experiments 2, 3, and 4 (all $ps > .05$); thus, we have collapsed across exposure corpora in reports of the analyses.

Experiment 2

If individuals are indeed sensitive to joint (rather than conditional) probabilities in action, we would expect such a sensitivity to enable adults to discover the segmental structure within the same exposure corpora used in Experiment 1. Thus, in Experiment 2, we employed the Experiment 1 exposure corpora, but participants were asked during test to report whether high-frequency actions were more familiar than low-frequency actions. Such a comparison allowed us to keep conditional probability constant (i.e., the conditional probabilities for elements in both high- and low-frequency actions were 1.0), while varying joint probability (co-occurrence frequency) information.

Method

Participants A total of 32 university students who had not participated in Experiment 1 (14 female, 18 male) received course credit for participating in the study.

Materials and procedure The same four exposure corpora were used as in Experiment 1, but different test sequences were selected. Specifically, we exhaustively paired every high-frequency action (120 tokens during exposure) with every low-frequency action (60 tokens during exposure), resulting in four test trials. We then reversed the order of presentation of each trial to create a total of eight test trials. As in Experiment 1, the order of the test trials was randomly determined, except that no pair could directly follow its reverse-order counterpart. Exposure and test were identical to those aspects of Experiment 1, except that the test trials now featured high- and low-frequency actions.

Results and discussion

Sensitivity to joint probability would be demonstrated by participants reporting high-frequency actions to be more familiar than low-frequency actions. Across the eight test trials, participants did indeed discriminate high-frequency actions from low-frequency actions, reporting high-frequency actions as more familiar, $M = 62.11\%$, $SD = 30.20$ (see Fig. 2) at levels significantly greater than would be predicted by chance, $t(31) = 2.27$, $p = .03$, Cohen's $d = 0.40$. Of the 32 participants, 21 chose the high-frequency action more frequently (five or more times out of eight trials) than the low-frequency action, which was not significant by a binomial test, $p = .10$. This nonsignificant result was unanticipated, but nevertheless the group performance compared against chance clearly demonstrated joint-probability sensitivity.

Together, the results from Experiments 1 and 2 suggest that while participants were sensitive to joint probabilities in sequences of human action, they were unable to segment action based on conditional probabilities. These findings are at odds with studies of statistical learning in other modalities, in which sensitivity to both joint and conditional probabilities has been demonstrated (e.g., Aslin et al., 1998; Fiser & Aslin, 2001, 2002a, 2002b; Graf Estes et al., 2007; but see Toro & Trobalón, 2005, in which a similar discrepancy was reported in rats—namely, demonstration of joint-probability but not conditional-probability learning in speech sounds).

What might explain the discrepancy between past studies of human statistical learning and our results? One possible answer lies in findings that learning of conditional probability requires more exposure to the statistically input than does learning of joint probability (e.g., Fiser & Aslin, 2002a; Graf Estes et al., 2007). Perhaps participants in Experiment 1 did not receive enough exposure to the input to allow for extraction of the higher-order conditional-probability statistics, suggesting that segmentation via conditional probability might be possible given more extensive exposure to the action stream.

Experiment 3

To address the possibility that participants in Experiment 1 had not received enough exposure to the action stream to extract conditional-probability statistics, in Experiment 3 we modified the procedure from Experiment 1 in one way: We constructed longer familiarization corpora. As in Experiment 1, we tested participants' discrimination of actions from frequency-balanced part actions, and successful discrimination would be indicative of conditional-probability learning.

Method

Participants Another 32 university students who had not participated in Experiment 1 or 2 (17 female, 15 male) received course credit for participating in the study.

Materials and procedure Four exposure corpora were created using the same actions as in Experiment 1, and the same part actions were also chosen for use during test. Actions that had occurred 120 times in Experiments 1 and 2 now occurred 180 times in Experiment 3, and actions that had occurred 60 times in Experiments 1 and 2 now occurred 90 times. The increase in action frequencies resulted in corpora that were approximately 35 min in length, an 11-min increase from Experiments 1 and 2.

In order to enact a procedure directly comparable to that used in Experiment 1 except for the change in length of the exposure corpora, we again presented participants with the same order of frequency-balanced action and part-action pairings used in Experiment 1. Experiment 3 procedurally was thus a direct replication of Experiment 1, and any success in action segmentation based on conditional probability would suggest that the null results seen in Experiment 1 were due to an insufficient amount of exposure to the action stream. Exposure and test were thus entirely identical to those aspects of Experiment 1, except that the exposure corpus was approximately 11 min longer.

Results and discussion

As in Experiment 1, learning of conditional probability would be indicated by above-chance selection of actions over part actions. The selections did not differ from chance levels, however. The mean action selection was 47.27% (SD = 27.08), $t(31)$ = −0.57, p = .57 (see Fig. 2). Only 11 participants selected actions more frequently (on five or more of eight test trials) than part actions, which did not significantly differ from chance levels according to a binomial test, p = .11.

The failure to find significant above-chance selection of actions in Experiment 3 appears to confirm the findings observed in Experiment 1, namely that adults are not able to segment human action according to conditional-probability information. Despite the additional exposure time, participants' selection of actions actually decreased (although this decrement was not a significant change from Experiment 1, $t(62)$ = 1.47, p = .15). Thus far, then, the assembled evidence across the present experiments indicates that adults display sensitivity only to lower-order joint-probability information in human action but do not learn conditional probabilities in action. Alternative explanations, however, should be considered. The average mean action selection exceeded 50% in Experiment 1, in which participants were tested on conditional-probability learning with a shorter, less attentionally taxing exposure time. Although the result was not significant, the fact that actions tended on average to be selected as more familiar than part actions hints at the possibility that at least some individuals were successfully computing conditional probabilities and responding on this basis. Moreover, it is possible that segmentation based on conditional probability might be demonstrated if we were to increase exposure time even more than was done in Experiment 3. However, given the marginal *decrease* in performance observed after our original decision to lengthen the exposure corpus, we designed Experiment 4 to investigate conditional-probability learn-ing in a way that did not require participants to watch longer sequences of actions.

Experiment 4

Experiment 4 included test pairings that allowed us to address conditional-probability learning in a new way. Recall that in Experiments 1 and 3, participants saw *only* frequency-balanced actions and part actions at test, and in the test phase of Experiment 2, they saw *only* pairings of actions with one another. This approach allowed for separate assessments of conditional probability and joint probability across experiments. In Experiment 4, however, we exhaustively paired all actions with part actions at test. This produced a subset of frequency-balanced pairings as well as pairings in which actions had appeared more frequently than part actions—that is, "frequency-unbal-anced" pairs. The logic behind this design was as follows: If segmentation in frequency-unbalanced trials is accomplished *only* on the basis of sensitivity to joint probability (the conclusion seemingly warranted by the data from Exps. 1, 2, and 3), performance on the unbalanced trials should not relate to performance on trials in which systematic action selection would require extraction of conditional probabilities (i.e., frequency-balanced trials). That is, a "joint-probability-only account" would predict that performance on frequency-unbalanced trials would be independent of performance on frequency-balanced trials. If, on the other hand, a relationship were to be found between performance on the two types of trials, this would suggest that joint probability sensitivity did *not* solely contribute to people's past successful detection of actions. To address these alternatives, we examined the relationship between performance on frequency-balanced and frequency-unbalanced trials.

The joint-probability-only account also entails a more specific prediction that is important for the present study: If indeed joint probability is the only statistic being computed by participants, individuals who tend to select actions at higher rates on frequency-unbalanced pairings should still be at chance when assessing action selection on frequency-balanced pairings. That is, if actions were selected based *only* on sensitivity to joint probability information in the frequency-unbalanced pairings, one would not expect to see similarly high rates of action selection on the frequency-balanced pairings, because joint probability could not be used to detect actions on the latter type of trial. This prediction is an important corollary to the general idea that a joint-probability-only account entails a lack of relationship on frequency-balanced and frequency-unbalanced trials, because it allows for a direct and unambiguous analysis of

conditional-probability learning by providing an opportunity to compare action selection against chance. Specifically, we planned an analysis to examine action selection against chance on frequency-balanced trials in individuals performing above versus below the median discrimination on frequency-unbalanced trials.

Method

Participants A total of 32 university students (20 female, 12 male) who had not participated in any prior studies of statistical learning received class credit for participation.

Materials and procedure We used the same 25-min exposure corpora used in Experiments 1 and 2. For test trials, we exhaustively paired every action with every part action, resulting in a total of 16 test trials with four frequency ratio differences; 4 trials apiece featured actions that were ten times, five times, and two times more frequent than the comparison part actions—hereafter referred to as the 10×, 5×, and 2× trials, respectively—as well as the 4 frequency-balanced pairs—hereafter referred to as the *equal* trials (see Table 3). We thus used the same frequency-balanced test trials employed in Experiment 1 (though we did not reverse the order of presentation), and we additionally featured frequency-unbalanced test trials in which actions were more frequent than their comparator part actions. Exposure and test were identical to those aspects of Experiments 1, 2 and 3, except that there were 16 test trials featuring both frequency-balanced and frequency-unbalanced pairs.

Results and discussion

Our first analysis examined participants' overall discrimination of actions as more familiar than part actions, independent of frequencies. A one-sample *t* test revealed that actions were selected at significantly above-chance levels ($M = 67.19\%$, $SD = 18.58$), $t(31) = 5.23$, $p < .001$, Cohen's $d = 0.93$. Of the 32 participants, 24 selected actions more frequently (on 9 or more of the 16 test trials) than part

Table 3 Small motion element co-occurrence frequency trial types in Experiment 4

Co-Occurrence Ratio of Action to Part Action	10×	5×	2×	Equal
Action frequency	120	60	120	60
Part-action frequency	12	12	60	60

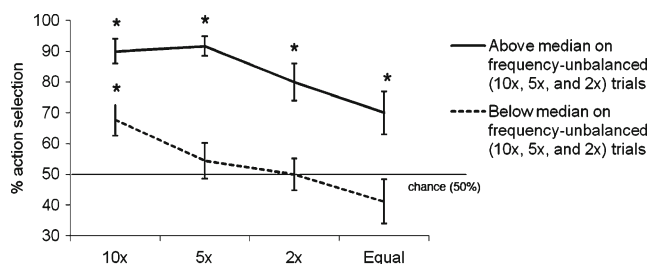actions, which was statistically significant by a binomial test, $p < .01$.

In order to evaluate whether there was any evidence of learning based on conditional probabilities, we next restricted analyses only to the four trials in which actions and part actions were equally frequent during exposure. Here, similar to the results from Experiments 1 and 3, we did not find any evidence of segmentation based on sensitivity to conditional probability; on the contrary, participants were unsystematic in their selection of actions ($M = 54.69\%$, $SD = 31.39$), $t(31) = 0.85$, $p = .41$. Also, only 14 participants chose actions more than half of the time (i.e., on more than two of the four test trials), which was not significant by a binomial test, $p = .60$.

We now turn to the primary analyses of this experiment, an examination of performance on frequency-balanced pairs in relation to performance on frequency-unbalanced pairs. A joint-probability-only account would predict no relationship, because action selection on frequency-unbalanced trials would be enabled *solely* by joint-probability learning, and recognition of actions on frequency-balanced trials would not benefit from this ability. In contrast, a positive relationship would suggest that participants' overall above-chance action selection was not due solely to joint-probability learning. We first examined whether average action selection on frequency-balanced trials was correlated with action selection on frequency-unbalanced trials. A Pearson correlation revealed a significant positive relationship, $r(30) = .47$, $p = .007$.

The fact that a positive correlation existed between action selection on frequency-balanced and frequency-unbalanced trials argues against the idea that people are only sensitive to joint probabilities in action. The presence of this relationship motivated the second analysis, in which we examined action selection on frequency-balanced trials of individuals who were above versus below the median on selections of actions in frequency-unbalanced trials. Again, the logic of our analysis was as follows: If action selection on frequency-unbalanced trials were due *only* to joint-probability learning, we would still expect to see at-chance performance in the frequency-balanced trials. This is not, however, what we observed. Instead, we obtained clear evidence for conditional-probability learning. The mean action selection in the equal (frequency-balanced) trials for individuals above the median on unbalanced trials was 70.0% ($SD = 27.06$), which was significantly greater than chance, $t(14) = 2.86$, $p = .01$, Cohen's $d = 0.74$ (see Fig. 3). Additionally, action selection was significantly above chance in all other pairings as well (an unsurprising result, given that we were selecting above-median performers on a task demonstrated to already feature overall greater-than-chance performance): In 10× trials, the mean action selection was 90.0% ($SD = 15.81$), $t(14) = 9.79$; in 5×

Fig. 3 Mean action selections in Experiment 4 by above-median and below-median individuals. The error bars represent ± 1 standard error. *Different from chance, *p* < .05

trials, selection was 91.67% (*SD* = 12.2), *t*(14) = 13.23; and in 2× trials, selection was 80.0% (*SD* = 25.53), *t*(14) = 4.94, all *p*s < .001 (see Fig. 3).[3]

In contrast, individuals scoring at or below the median on frequency-unbalanced trials did *not* show above-chance action selection in the equal trials, *M* = 41.18% (*SD* = 21.23); this performance did not differ from chance levels, *t* (16) = −1.24, *p* = .23 (see Fig. 3). As well, above-chance action selection in the below-median individuals was only seen when actions were ten times more frequent than part actions (*M* = 67.65%, *SD* = 21.22), *t*(16) = 3.43, *p* = .003. Action selection on 5× trials (*M* = 54.41%, *SD* = 23.78) and

2× trials (*M* = 50.0%, *SD* = 21.65) did not differ from chance levels, *t*(16) = 0.77 and 0, respectively; *p*s > .05 (see Fig. 3).

Taken together, these results suggest that a certain subset of individuals—namely, those selecting actions at high rates in the frequency-unbalanced trials—were in fact sensitive to conditional probability. The overall significant selection of actions over part actions was thus *not* due solely to joint-probability learning;[4] the findings clearly indicate that conditional-probability learning is achieved by some individuals.

General discussion

The present experiments addressed what types of statistics people can extract from a stream of dynamic human action, with a specific focus on two distinct types of statistical information—namely, joint and conditional probability. Whereas joint probability can be calculated via sensitivity to co-occurrence frequencies of multiple events, conditional-probability learning in our study required extraction of predictive relationships among multiple dynamic human actions. We found positive evidence for both types of learning, although conditional-probability learning was seen in only a subset of our participants.

Notably, we did not obtain evidence for conditional-probability learning from the traditional experimental

---

[3] A possible concern is that individuals' familiarity judgments on early test trials influenced later judgments—that is, earlier test trials might bias selection of sequences on later test trials. A sequence judged as more familiar on an early test trial would, for instance, plausibly be more likely to be judged as familiar on a later test trial, due only to the recent judgment on the earlier trial, and not to statistical information computed during exposure. This possibility allows for an alternative explanation for the above-chance performance on frequency-balanced trials for above-median individuals. The specific version of this explanation that would be most problematic for the present account is the following scenario: On a given early test trial, an individual relies on joint-probability calculation to judge that a low-frequency action (*f* = 60) is more familiar than a low-frequency part action (*f* = 12) (e.g., for Corpus 1, judging the low-frequency action *drink/twirl/read* as more familiar than the lower-frequency part action *read/rattle/slide*). On a later test trial, when this same low-frequency action is pitted against an equal-frequency part action (*f* = 60) (e.g., from Corpus 1, *clean/under/feel*), the individual is more likely to select the action as opposed to its frequency-balanced comparator due to its having recently been selected on the earlier trial, *not* to its higher conditional probability. To ensure that this could not explain the above-chance frequency-balanced trial performance in the present study, we ran a secondary analysis examining performance only on those frequency-balanced trials that featured an action that had not yet been paired with a low-frequency (*f* = 12) part action. That is, we examined performance in above-median individuals on frequency-balanced trials on which an individual would not yet be biased to favor a given action due to its having been selected earlier, based on its higher frequency relative to its comparator. We again found above-chance performance on this restricted set of frequency-balanced trials (*M* = 80.95, *SD* = 33.88), *t*(14) = 3.54, *p* = .003. Thus, the likelihood that this "spillover" alternative explanation could account for our results is substantially reduced.

[4] Figure 3 displays decreasing performance as the frequency of actions relative to part actions decreases, hinting at a role for frequency in contributing to the findings in Experiment 4. There were indeed differences across trial types (10x, 5x, 2x, and equal) in above-median individuals, *F*(3, 42) = 4.14, *p* = .01, as well as below-median individuals, *F*(3, 48) = 3.22, *p* = .03. Contrasts revealed significant linear trends for both the above-median, *F*(1, 14) = 7.93, *p* = .01, and the below-median, *F*(1, 16) = 8.59, *p* = .01, individuals, with rates of action selection decreasing as the ratios of action:part-action frequency decreased. While it might thus appear that the observed linear trends are due to decreasing action frequencies relative to the comparator part actions, this conclusion is not definitively warranted. The conditional-probability statistics were not equal for all part actions used in the test trials. Specifically, for pairings in 10x and 5x trials, conditional-probability statistics for the part actions were on average lower than those for the part actions in 2x and equal trials. Using part actions from Corpus 1 as an example, the part actions *clean/under/feel* and *blow/look/empty* (used in 2x and equal trials) on average would have higher conditional probabilities than *read/rattle/slide* and *poke/drink/twirl* (used in 10x and 5x trials), because the likelihood of *under* being followed by *feel*, and of *look* being followed by *empty*, was .5, whereas the likelihood of *read* being followed by *rattle*, and of *poke* being followed by *drink*, was .2.

The appropriate test of the role of joint probability assesses learning in paired comparisons between 10x and 5x, as well as 2x and equal, trials, because conditional probability statistics were constant within the 10x/5x and 2x/equal groupings. Here, though, paired-samples *t* tests revealed nonsignificant differences [for 10 x vs. 5x, *t* (31) = 1.16, *p* = .26; for 2x vs. equal, *t*(31) = 1.68, *p* = .10], likely due in part to decreased power. Thus, data from Experiment 4 cannot be used to make strong claims regarding the independent role of varying joint probability in action segmentation.

designs used in the past to reveal learning involving other types of input (e.g., speech, static shapes, and shape configurations; Aslin et al. 1998; Fiser & Aslin, 2001, 2002a). Namely, we first attempted to demonstrate conditional-probability learning by asking whether an entire group of participants could discriminate at above-chance levels between action sequences equated in terms of joint probability but differing in terms of conditional probability, the standard method used by others in the past. Null results for this experiment (Exp. 1), combined with positive results for a similar assessment of joint-probability learning (Exp. 2), appeared to suggest that past findings of statistical learning of actions (e.g., Baldwin et al. 2008) were due solely to individuals' tracking of the co-occurrence frequencies of action elements.

However, alternative explanations were available; past findings of conditional-probability learning have typically demonstrated that it requires longer exposure for successful learning relative to the exposure required for joint-probability learning (e.g., Fiser & Aslin, 2002a; Graf Estes et al. 2007), and thus we addressed this possibility in two ways. First, we exposed participants to a longer corpus of actions, with the expectation that this would aid participants in the relatively difficult task of extracting conditional-probability information. Despite the provision of this additional information, however, participants actually selected actions at rates even lower than before (although this decrement was not statistically significant). Thus, we conducted a final experiment in which we were able to examine conditional-probability learning in a different and novel way.

In this last experiment, we provided participants with discrimination (test) trials that varied in terms of the frequencies of actions relative to their comparator part actions; some test trials featured actions that occurred more frequently than their paired part action, whereas others featured actions that had occurred equally as frequently. This design allowed us to examine the relationship between performance on frequency-unbalanced trials with performance on frequency-balanced trials. Performance on the two types of trials was highly correlated. Further, individuals who tended to select actions over part actions on frequency-unbalanced trials (at above-median levels) *also* selected actions over part actions on frequency-balanced trials. Action selection among these individuals was significantly above chance, demonstrating that at least this subset of individuals were capable of tracking conditional probability in dynamic human action.

Our findings are in many ways consistent with past studies of statistical learning in other domains. Extensive past research in statistical learning of many types of information, including speech sounds and simple static shapes, suggested that humans (both infants and adults) are capable of calculating the conditional probability expressing predictive relationships among elements (Aslin et al. 1998; Aslin et al. 2001; Fiser & Aslin, 2001, 2002a, 2002b; Graf Estes et al. 2007). We similarly demonstrated that extraction of higher-order conditional-probability statistics is possible within the action domain. However, our results also indicated that in the action-processing context, only a subset of our participants demonstrated sensitivity to conditional probability. This result stands in stark contrast to past studies of conditional-probability learning, which have demonstrated such learning on a group level (i.e., with entire samples).

An important question thus arises from our findings: Why was conditional-probability learning of action only observed in a subset of our participants? This is perhaps especially puzzling given that we exposed participants to a number of actions that actually exceeded the number of words used in the language-learning study from which we adapted our design. For instance, whereas Aslin et al. (1998) used an exposure that contained 270 words, our adult participants saw either 360 actions across the course of exposure (Exps. 1, 2, and 4) or 540 actions (Exp. 3). Despite this increase in the sheer number of statistically structured units, however, our participants on a group level failed to show segmentation with either the short or the long exposure.

One possible explanation for the discrepant findings is that our small motion elements may have differed in encodability in comparison to the syllables used by Aslin et al. (1998), as well as the simple shapes used in past studies of visual statistical learning (Fiser & Aslin, 2001, 2002a, 2002b). Recall that Gebhart et al. (2009) found that participants required much more exposure to nonspeech noises to learn the underlying statistical structure, and that these researchers attributed this difference to the fact that the nonspeech noises were less familiar, and thus likely less encodable than speech syllables. Although the SMEs that we used in the present experiments are likely at least moderately familiar, other differences exist between our stimuli and the elements used in past studies that may similarly contribute to encodability. First, in terms of basic perceptual features, our SMEs are arguably far more complex than the units used in past studies of visual statistical learning. In order to encode our SMEs, participants would have had to process evanescent, dynamic events rather than static simple shapes or shape configurations. Second, our elements possessed meaning in and of themselves that both speech syllables and shapes did not. For example, the SME *rattle* features an event (albeit brief) that itself invites sophisticated and potentially processing-intensive inferences regarding the intentions of the actor. Further, our elements were nameable (e.g., *rattle*, *blow*, *drink*), whereas the syllables and shapes used in past studies

had no conventional linguistic labels (since the shapes used in past studies were relatively simple but had no linguistic labels). It is possible that the more complex perceptual attributes of our SMEs made them more difficult to encode. Further, although the richer conceptual and linguistic content inherent in our SMEs might make them *individually* more meaningful, and hence possibly more memorable, it is possible that integrating them into higher-level action units was consequently more difficult. Determining what attributes contribute to encodability (e.g., perceptual complexity, conceptual richness, or linguistic factors) is an inviting topic for future research in statistical learning and can further contribute to resolving whether differences observed in statistical learning are due to mechanism differences per se, as opposed to stimulus-based encodability differences.

Assuming that action may indeed be harder to encode, our findings give rise to yet another unanswered question: What contributed to only a subset of our participants succeeding in exploiting conditional probability for segmentation purposes? Is it possible that certain individuals would *never* be capable of learning predictive relationships among action elements on such a higher-order level, and that the below-chance performance on frequency-balanced trials demonstrated by half of our participants is indicative of this lack of ability? Given the likely importance of calculating such statistics (e.g., learning to predict events based on the contingent relationships among action elements), this seems an improbable conclusion. Indeed, this ability seems especially fundamental with respect to the role it could play in segmentation, according to theories of action processing such as Zacks and colleagues' event segmentation theory (Kurby & Zacks, 2008; Zacks et al. 2007). Recall that knowledge of sequential dependencies gained through statistical learning is posited as one source of information that feeds into the predictive system responsible for creating and maintaining the stability of the event model. If some individuals truly were limited to learning only about co-occurrence frequencies of actions, as opposed to contingent, predictive relationships among elements, the power of statistical learning in contributing to their action processing would be substantially reduced. Further, we would then likely see profound downstream variability in the way that different individuals process, interpret, and predict actions, and yet the consensus from the literature is that both segmentation and higher-level mental-state inferences unfold relatively uniformly and automatically, at least among normally developing individuals (e.g., Wellman, 2002; Zacks et al. 2001).

Instead, it seems likely that external transient factors contributed to our individuals' varying performance profiles, including variations in motivation or alertness. It is also possible that these factors may have included more stable individual differences in functions such as working memory or allocation of attentional resources, basic cognitive processes that have been demonstrated to vary among individuals (see, e.g., Baddeley, 2001). Determining exactly which factors were at play in contributing to these differences is an important topic for future work. In general, most research on statistical learning has focused on comparing learning of various types of statistics (e.g., joint vs. conditional probability or adjacent vs. nonadjacent dependencies; see, e.g., Aslin et al. 1998; Gebhart et al. 2009; Newport & Aslin, 2004; Toro & Trobalón, 2005) or differences across various modalities (e.g., Conway & Christiansen, 2005) rather than differences among individuals. However, some studies have been directed at revealing variation in performance on a single statistical learning task; for example, Ludden and Gupta (2000) showed that statistical learning of speech is impaired when cognitive load demands are increased, hinting at the possibility that stable individual differences related to cognitive processing might relate to statistical learning. Further, Evans, Saffran, and Robe-Torres (2009) found that statistical learning of both speech and nonspeech stimuli was impaired in children with specific language impairment and that individual differences in learning were correlated with vocabulary. Variation has also been demonstrated in normally developing populations; for instance, Misyak, Christiansen, and Tomblin (2010) found that differences in individuals' learning trajectories of nonadjacent dependencies in linguistic stimuli predicted later performance on an online language-processing task. These observations, coupled with our own results, point to a clear need to further elucidate the causes for these variations as well as their outcomes.

Another way of gaining understanding regarding individual variation—as well as more broadly in elucidating the processes underlying segmentation—would be to explore alternative measures of segmentation. We chose to assess participants' explicit reports of familiarity for different sequences, a methodological decision that allowed us to draw comparisons to a number of past studies in other domains that used the same measure (e.g., Fiser & Aslin, 2001, 2002a). However, it would be instructive to examine segmentation evidenced in other ways, as well. For instance, Abla and Okanoya (2009) studied the event-related potentials of individuals as they watched a statistically structured sequences of shapes. The researchers found that individuals who later performed well on a familiarity-based behavioral segmentation test also displayed larger N400 amplitudes at the onsets of statistically coherent shape triplets after a period of exposure. The results of this study hold promise in providing a more implicit, online measure of segmentation, and future work in the action-processing domain might benefit from a similar incorporation of neurophysiological measures. In another exploration

of alternative methods for assessing segmentation, we are currently adapting a methodology devised by Hard, Recchia, and Tversky (2011), who discovered evidence of attentional surges in response to action boundaries. The prediction derived from these findings, and the one that is currently being explored, is that individuals' attention at the junctures between actions should be modulated as a function of statistical learning. Specifically, as an observer learns the statistical structure characterizing a sequence of actions, they should start to show similar surges of attention at action boundaries (i.e., at onsets of SME triplets comprising actions). This method may shed light on both the dynamic process of statistical learning as it unfolds, and also provide further insight into possible mediating attentional processes underlying the individual differences that we observed.

In sum, our results point to important similarities as well as dissimilarities in the nature of statistical learning. People can indeed learn the conditional probabilities structuring dynamic human action, a result that parallels those in other domains. Our results thus indicate that prior findings of statistical learning in actions (Baldwin et al. 2008) are likely due, at least in part, to individuals' sensitivity to conditional probability. On the other hand, we also demonstrated that there is substantial variation in the ability to detect action segments on the basis of conditional probability; only individuals who were also especially successful in detecting more-frequent actions showed sensitivity to conditional probability. The ability to recognize statistical regularities in action is likely a crucial component of the human action-processing system, allowing observers to use bottom-up information to feed predictions about how events will unfold. Our findings regarding how joint-probability learning and conditional-probability learning contribute to this process mark an important step in understanding the function that statistical learning has in people's ability to process and make sense of human action.

References

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research*, 64, 185–190. doi:10.1016/j.neures.2009.02.013

Aslin, R.N., Saffran, J.R., Newport, E.L. (1998). Computation of conditional probability statistics by human infants. *Psychological Science*, 9, 321–324. doi:10.1111/1467-9280.00063

Aslin, R. N., Slemmer, J. A., Kirkham, N. Z., & Johnson, S. P. (2001). *Statistical learning of visual shape sequences*. Minneapolis, MN: Paper presented at the meeting of the Society for Research in Child Development.

Avrahami, J., Kareev, Y. (1994). The emergence of events. *Cognition*, 53, 239–261. doi:10.1016/0010-0277(94)90050-7

Baddeley, A. D. (2001). Is working memory still working? *American Psychologist, 56,* 849–864.

Baird, J. A., & Baldwin, D. A. (2001). Making sense of human behavior: Action parsing and intentional inference. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 193–206). Cambridge, MA: MIT Press.

Baldwin, D.A., Andersson, A., Saffran, J.R., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106, 1382–1407. doi:10.1016/j.cognition.2007.07.005

Baldwin, D.A., & Baird, J.A. (2001). Discerning intentions in dynamic human action. *Trends in Cognitive Sciences*, 5, 171–178. doi:10.1016/S1364-6613(00)01615-6

Baldwin, D.A., Baird, J.A., Saylor, M.M., & Clark, M.A. (2001). Infants parse dynamic action. *Child Development*, 72, 708–717. doi:10.1111/1467-8624.00310

Conway, C.M., & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 24–39. doi:10.1016/S0278-7393(05)80004-7

Evans, J.L., Saffran, J.R., & Robe-Torres, K. (2009). Statistical learning in children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 52, 321–335. doi:10.1044/1092-4388(2009/07-0189)

Fiser, J., & Aslin, R.N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12, 499–504. doi:10.1111/1467-9280.00392

Fiser, J., & Aslin, R.N. (2002a). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 458–467. doi:10.1037//0278-7393.28.3.458

Fiser, J., & Aslin, R.N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99, 15822–15826. doi:10.1073/pnas.232472899

Gebhart, A.L., Newport, E.L., & Aslin, R.N. (2009). Statistical learning of adjacent and non-adjacent dependencies among non-linguistic sounds. *Psychonomic Bulletin & Review*, 16, 486–490. doi:10.3758/PBR.16.3.486

Graf Estes, K., Evans, J.L., Alibali, M.W., & Saffran, J.R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, 18, 254–260. doi:10.1111/j.1467-9280.2007.01885.x

Hard, B.M., Recchia, G., & Tversky, B. (2011). *The shape of action.* Manuscript submitted for publication

Hard, B. M., Tversky, B., & Lang, D. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition, 34,* 1221–1235.

Hunt, R. H., & Aslin, R. N. (1998). Statistical learning of visuomotor sequences: Implicit acquisition of sub-patterns. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the twentieth annual conference of the cognitive science society* (pp. 508–513). Mahwah, NJ: Erlbaum.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S.P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition*, 83, B35–B42. doi:10.1016/S0010-0277(02)00004-5

Kurby, C.A., & Zacks, J.M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12, 72–79. doi:10.1016/j.tics.2007.11.004

Ludden, D., & Gupta, P. (2000). Zen in the art of language acquisition: Statistical learning and the Less is More hypothesis. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 812–817). Mahwah, NJ: Erlbaum.

Marcus, G.F., Fernandes, K.J., & Johnson, S.P. (2007). Infant rule learning facilitated by speech. *Psychological Science*, 18, 387–391. doi:10.1111/j.1467-9280.2007.01910.x

Misyak, J.B., Christiansen, M.H., & Tomblin, J.B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology, 1*, 1–9. doi:10.3389/fpsyg.2010.00031

Newport, E.L., & Aslin, R.N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162. doi:10.1016/S0010-0285(03)00128-2

Newtson, D., & Enquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*, 435–450. doi:10.1016/0022-1031(76)90076-7

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35*, 847–862. doi:10.1037/0022-3514.35.12.847

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning* (pp. 64–99). New York: Appleton-Century-Croft.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996a). Statistical learning by 8-month old infants. *Science, 274*, 1926–1928. doi:10.1126/science.274.5294.1926

Saffran, J.R., Newport, E.L., & Aslin, R.N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621. doi:10.1006/jmla.1996.0032

Saffran, J.R., Johnson, E.K., Aslin, R.N., & Newport, E.L. (1999). Statistical learning of tone sequences by adults and infants. *Cognition, 70*, 27–52. doi:10.1016/S0010-0277(98)00075-4

Saffran, J.R., Pollak, S.D., Seibel, R.L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105*, 669–680. doi:10.1016/j.cognition.2006.11.00

Saylor, M., & Baldwin, D. A. (2004). Action analysis and change blindness: Possible links. In D. T. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 37–56). Westport, CT: Greenwood/Praeger.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Speer, N. K., Swallow, K. M., & Zacks, J. M. (2003). Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience, 3*, 335–345.

Swallow, K.M., & Zacks, J.M. (2008). Sequences learned without awareness can orient attention during the perception of human activity. *Psychonomic Bulletin & Review, 15*, 116–122. doi:10.3758/PBR.15.1.116

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

Tversky, B., Zacks, J. M., & Hard, B. (2008). The structure of experience. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: From perception to action* (pp. 436–464). New York: Oxford University Press.

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167–187). Oxford: Blackwell.

Zacks, J.M. (2004). Using movement and intentions to understand simple events. *Cognitive Science, 28*, 979–1008. doi:10.1016/j.cogsci.2004.06.003

Zacks, J.M., & Swallow, K.M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*, 80–84. doi:10.1111/j.1467-8721.2007.00480.x

Zacks, J.M., Speer, N.K., Swallow, K.M., Braver, T.S., & Reynolds, J.R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin, 133*, 273–293. doi:10.1037/0033-2909.133.2.273

Zacks, J.M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*, 29–58. doi:10.1037/0096-3445.130.1.29

Zacks, J.M., Kumar, S., & Abrams, R.A., Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition, 112*, 201–21.doi:10.1016/j.cognition.2009.03.007

Author Note