Learning from Actions and their Consequences: Inferring Causal Variables from Continuous Sequences of Human Action

Daphna Buchsbaum (daphnab@berkeley.edu)
Thomas L. Griffiths (Tom_Griffiths@berkeley.edu)
Alison Gopnik (gopnik@berkeley.edu)
Department of Psychology, University of California, Berkeley, Berkeley, CA 94720 USA

Dare Baldwin (baldwin@uoregon.edu)
Department of Psychology, 1227 University of Oregon, Eugene, OR 97403 USA

Abstract

In the real world causal variables do not come pre-identified or occur in isolation, but instead are imbedded within a continuous temporal stream of events. A challenge faced by both human learners and machine learning algorithms is identifying subsequences that correspond to the appropriate variables for causal inference. A specific instance of this problem is action segmentation: dividing a sequence of observed behavior into meaningful actions, and determining which of those actions lead to effects in the world. Here we present two experiments investigating human action segmentation and causal inference, as well as a Bayesian analysis of how statistical and causal cues to segmentation should optimally be combined. We find that both adults and our model are sensitive to statistical regularities and causal structure in continuous action, and are able to combine these sources of information in order to correctly infer both causal relationships and segmentation boundaries.

Keywords: Statistical learning; Causal inference; Action segmentation; Rational analysis; Bayesian inference

Introduction

Human social reasoning ability depends on understanding the relationship between actions, goals and outcomes. Learners must take a continuous stream of observed behavior and divide it into distinct meaningful actions. Determining which subsequences go together, and what outcomes they produce, is also an important instance of the more general problem of causal variable discovery (a similar problem – determining how spatially distributed observations should be encoded as variables – is discussed by Goodman, Mansinghka, & Tenenbaum, 2007). Prior research has shown that adults are able to segment videos of common everyday behaviors into coherent actions, corresponding to the goals and intentions underlying the actor's behavior (for a recent review see Kurby & Zacks, 2008), and that even young infants are sensitive to the boundaries between intentional action segments (Baldwin, Baird, Saylor, & Clark, 2001; Saylor, Baldwin, Baird, & LaBounty, 2007). However, little is yet known about both the types of information people use to detect the boundaries between actions, and the computations that allow us to predict and extract individual actions.

One potentially important source of information is statistical regularities in the action stream. There is now a large body of evidence suggesting that both infants and adults can use statistical patterns in spoken language to help solve the related problem of segmenting words from continuous speech (for a partial review, see Gómez & Gerken, 2000). Recently, Baldwin, Andersson, Saffran, and Meyer (2008) demonstrated that a similar sensitivity to statistical regularities in continuous action sequences may play an important role in action processing. However, a key difference between action segmentation and word segmentation is that intentional actions usually have effects in the world. In fact, many of the causal relationships we experience result from our own and others' actions, suggesting that understanding action may bootstrap learning about causation, and vice versa. Though recent work has demonstrated that both children and adults can infer causal relationships from conditional probabilities (Gopnik et al., 2004; Griffiths & Tenenbaum, 2005), the extent to which action understanding and causal learning mechanisms inform each other has yet to be explored. Here we present a combination of experimental and computational approaches investigating how the ability to segment action and to infer its causal structure functions and develops.

We first introduce a Bayesian analysis of action segmentation and causal inference, which provides an account of how statistical and causal cues to segmentation should optimally be combined. Next, we present two experiments investigating how people use statistical and causal cues to action structure. Our first experiment demonstrates that adults are able to segment out statistically determined actions, and experience them as coherent, meaningful and most importantly, causal sequences. Our second experiment shows that adults are able to extract the correct causal variables from within a longer action sequence, and that they find causal sequences to be more coherent and meaningful than other sequences with equivalent statistical structure. Finally, we look at the action segmentations and causal structures our Bayesian rational model predicts, when given the same experimental stimuli as our human participants. We conclude by discussing our results in the context of broader work, as well as its implications for more generalized human statistical learning abilities.

Bayesian Analysis of Action Segmentation

We created a Bayesian rational learner model that jointly infers action segmentation and causal structure, using statistical regularities and temporal cues to causal relationships in an action stream. This model provides us with a way to begin characterizing both the kinds of information available in the action stream, and what an optimal computational level solution to these inference problems might look like. To the extent that our model accurately reflects human performance, it provides additional support for the idea that people may similarly be

combining statistical and causal cues in their own inference.

We adapted the nonparametric Bayesian word segmentation model first used by Goldwater, Griffiths, and Johnson (2006) to the action domain, and also extended this model to incorporate causal information. Like the original word segmentation model, our model is based on a *Dirichlet process* (Ferguson, 1973), with actions composed of individual small motion elements taking the place of words composed of phonemes. In addition, we incorporated cause and effect information into the generative model, allowing some actions to be probabilistic causes. We describe this model in more detail in the following sections.

Generative Model for Action Sequences

Just as a sentence is composed of words, which are in turn composed of phonemes, in our model an action sequence A is composed of actions a_i which are themselves composed of motion elements m_j. We assume a finite set of possible actions, and that complete actions are chosen one at a time from this set, and then added to the the sequence. The conditional probability of the next action in the sequence $p(a_i|a_1...a_{i-1})$, is given by a standard algorithm known as the *Chinese Restaurant Process* (CRP). In the CRP customers enter a restaurant, and are seated at tables, each of which has an associated label. In this case, the labels are actions. When the i^{th} customer enters the restaurant, they sit at a table z_i, which is either a new table or an already occupied table. The label at table z_i becomes the i^{th} action in our sequence with

$$p(z_i = k|z_1...z_{i-1}) = \begin{cases} \frac{n_k}{i-1+\alpha_0}, & 0 \leq k \leq K \\ \frac{\alpha_0}{i-1+\alpha_0}, & k = K+1 \end{cases} \quad (1)$$

where n_k is the number of customers already at table k, and K is the number of previously occupied tables. So, the probability of the i^{th} customer sitting at an already occupied table depends on the proportion of customers already at that table, while the probability of starting a new table depends on the *concentration parameter* α_0.

Whenever a customer starts a new table, an action a_k must be associated with this table. Since multiple tables may be labeled with the same action, the probability that the next action in the sequence will have a particular value $a_i = w$ is

$$p(a_i = w|a_1...a_{i-1}) = \frac{n_w}{i-1+\alpha_0} + \frac{\alpha_0 P_0(a_i = w)}{i-1+\alpha_0} \quad (2)$$

where n_w is the number of customers already seated at tables labeled with action w. In other words, the probability of a particular action $a_i = w$ being selected is based on the number of times it has already been selected, and the probability of generating it anew. We draw new action labels from the *base distribution* P_0. Actions are created by adding motions one at a time, so that $P_0(a_i = w)$ is simply the product of action w's component motion probabilities, with an added assumption that action lengths are geometrically distributed with

$$P_0(a_i) = p_\#(1-p_\#)^{n-1} \prod_{j=1}^{n} p(m_j) \quad (3)$$

where n is the length of a_i in motions, $p_\#$ is the probability of ending the action after each motion, and $p(m_j)$ is the probability of an individual motion. Currently, we use a uniform probability over all motions. We assume that action sequence length is also geometrically distributed, so we use the same equation for the overall sequence probability, substituting actions for motions, and $p_\$$ (the probability of ending the sequence A after the current action) for $p_\#$. In this work $p_\# = 0.95$, which represents a bias towards finding smaller length actions, $p_\$ = 0.001$ biases the model towards sequences made up of more actions, and $\alpha_0 = 3$ represents an expectation that the set of all possible actions is relatively small.

Generative Model for Events

The action sequence A also contains non-action events e, which can occur between motions. In our model, some actions are causal sequences, and are followed by an event with high probability. Each unique possible action a_w has an associated binary variable $c_w \in \{0,1\}$ that determines whether or not the action is causal with $c_w \sim \text{Bernoulli}(\pi)$. If an action is a causal sequence, then it is followed by an event with probability ω. We use a small fixed value ε for the probability of an effect occurring after a non-causal sequence (in the middle of an action, or after a non-causal action. See Figure 1). For this work, we used $\varepsilon = 0.00001$ and $\omega = 0.999$, which represent our assumption that events are very unlikely to follow non-causal sequences, and very likely to occur after actions that are causal sequences. We used $\pi = 0.05$, which represents an assumption that relatively few actions are causes for a particular effect.

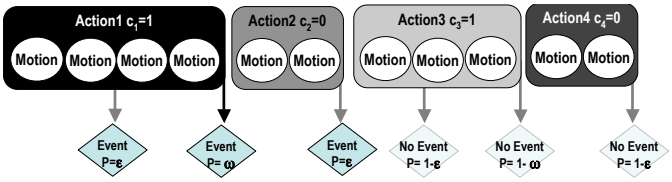

Figure 1: A theoretical action sequence depicting causal relationships in the model.

Inferring Segmentation and Causal Structure

An unsegmented action sequence consists of the motions m_j without any breaks between them. Given such a sequence, how do we find the boundaries between actions? A segmentation hypothesis h indicates whether there is an action boundary after each motion m_j. For a given segmentation hypothesis h, and unsegmented action sequence d, we use Bayes rule $p(h|d) \propto p(d|h)p(h)$ to infer the posterior distribution $p(h|d)$.

To do this, we can use a standard Markov chain Monte Carlo method known as Gibbs sampling (Gilks, Richardson, & Spiegelhalter, 1996). The key property of a Gibbs sampler is that it converges to the posterior distribution, allowing us to sample segmentation hypotheses from $p(h|d)$ (a more detailed explanation of using Gibbs sampling to infer a segmentation is given in Goldwater et al., 2006). We can also use Gibbs sampling to infer the posterior distribution over causal relationships between actions and events. In this case

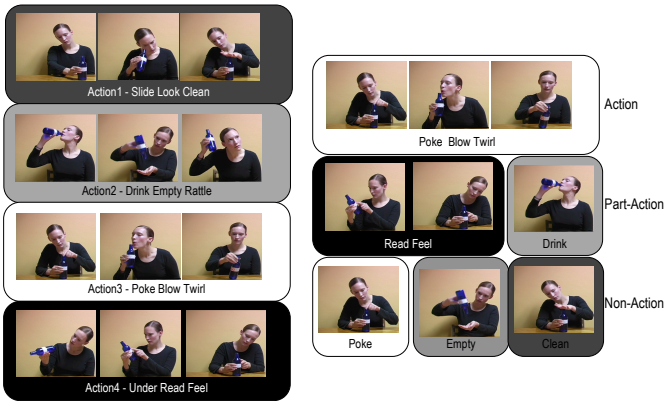

Figure 2: Left: Four actions composed of three unique motions each were used to create the Experiment 1 exposure corpus. Right: Example Action, Part-Action and Non-Action.

a causal structure hypothesis h consists of values c_w for all the actions found in the inferred segmentation.

Predictions for Human Segmentation

An important feature of this model is that action segmentation and causal structure are learned simultaneously, and interdependently. At each iteration, the inferred actions help determine the inferred causal structure and vice versa, because the two are linked in the generative model. This corresponds to our prediction that people also believe actions and causal effects go hand in hand. If statistical action structure is also a cue to causal relationships then, like our model, adults should think statistically grouped actions are more likely to be potential causes than other equivalent sequences. This prediction is tested in Experiment 1. Second, if people believe that causal sequences of motion are also likely to be actions, then adults should be able to not only segment out causal sequences, but find them to be more meaningful and coherent than other sequences with equivalent statistical regularities. This prediction is tested in Experiment 2.

Experiment 1: Using Statistical Cues

The structure of this experiment is similar to that used in previous action segmentation experiments (Baldwin et al., 2008; Meyer & Baldwin, 2008). Since previous work has established that adults are able to recognize artificial *actions* grouped only by their statistical regularities, we wanted to investigate whether these groupings are also considered meaningful, and whether they are inferred to be causal. Specifically, we hypothesized that participants would judge these artificial *actions* to be more coherent and meaningful than similar *non-action* and *part-action* sequences (see Figure 2), and would also view *actions* as more likely to generate a (hidden) effect than *non-actions* and *part-actions* .

Method

Participants Participants were 100 U.C Berkeley undergraduate students, who received course credit for participating. Participants were randomly assigned to view one of the two exposure corpora, and were also randomly assigned to one of three follow-up question conditions. All participants were instructed to attend closely to the exposure corpus, and were told that they would be asked questions about it later. Thirty participants were assigned to each of the first two conditions and 40 participants were assigned to the last condition.

Stimuli Similar to Baldwin et al. (2008), we used 12 individual video clips of object-directed motions (referred to as *small motion elements* or SMEs in the previous work), to create four *actions* composed of three SMEs each (see Figure 2). The SMEs in this experiment are identical to those in Meyer and Baldwin (2008). As in previous work, SMEs were sped up slightly and transitions were smoothed using iMovie HD, to make the exposure corpus appear more continuous.

We created a 25 minute exposure corpus by randomly choosing actions to add to the sequence, with the condition that no action follow itself, and that all actions and transitions between actions appear an equal number of times, resulting in 90 appearances of each action and 30 appearances of each transition. We also created four *non-action* and four *part-action* comparison stimuli, where a non-action is a combination of three SMEs that never appear together in the exposure corpus, and a part-action is a combination of three SMEs that appears across a transition (e.g. the last two SMEs from the first action and the first SME from the second action, see Figure 2). Finally, to ensure that none of our randomly assembled actions were inherently more causal or meaningful, we created a second exposure corpus, using the non-action SME combinations of the first corpus as the actions of the second corpus.

Procedure Following the exposure corpus, participants in the *familiarity condition* were presented with all 12 actions, non-actions and part-actions individually, and asked "How familiar is this action sequence?". They responded by choosing a value on a 1 to 7 Likert scale, with 1 representing "not familiar" and 7 representing "very familiar" (other than the use of ratings instead of a forced choice format, this condition is almost identical to Baldwin et al., 2008). In the *causal condition*, participants were given a "hidden effect" cover story before viewing the exposure corpus. These participants were told that certain actions would cause the bottle being manipulated to play music, but that they would be watching the video with the sound off. Following the exposure corpus, these participants were asked "How likely is this sequence to make the bottle play a musical sound?", with 1 representing "not likely" and 7 representing "most likely". Finally, in the *coherence condition*, participants were asked the question "how well does this action sequence go together?". They were given the example of removing a pen cap and then writing with the pen as "going together" and of removing a pen cap and then tying your shoes as "not going together". They then rated all test items on a scale with 1 being "does not go together" and 7 being "goes together well".

For all conditions, we used a custom Java program to present video of action sequences and collect ratings.

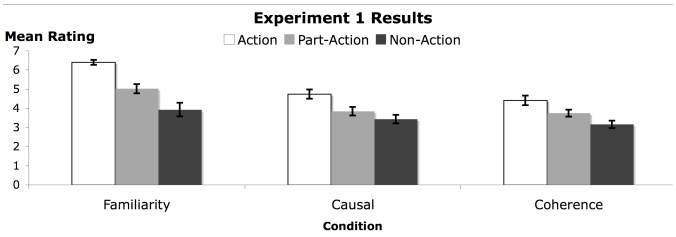

Figure 3: Results of Experiment 1. Error bars show one standard error.

The program presented all 12 actions, non-actions and part-actions individually and in a random order.

Results

We analyzed all results using 2×3 ANOVAs on exposure corpus (1 or 2) and sequence type (action, non-action, part-action). No effects of exposure corpus were found.

Ratings from 27 participants in the *familiarity condition* were analyzed (data from three additional participants who rated all sequences identically as either a 1 or 7 was discarded). As predicted by previous results (Baldwin et al., 2008; Meyer & Baldwin, 2008), there was an overall significant effect of sequence type $F(2, 50)= 25.14$, $MSE= 41.12$, $p < 0.0001$, with actions rated significantly more familiar than part-actions and non-actions $t(26)= 5.84$, $p < 0.0001$, one sample t-test on contrast values, and part-actions rated significantly more familiar than non-actions $t(26)= 3.65$, $p < 0.002$.

Ratings from 29 participants in the *causal condition* were analyzed (data from one additional participant was discarded). As predicted, there was an overall significant effect of sequence type $F(2,54)= 10.20$, $MSE= 12.869$, $p < 0.0002$, with actions rated as significantly more likely to cause a musical effect than part-actions or non-actions $t(28)= 2.36$, $p < 0.01$, one sample t-test on contrast values, and part-actions rated significantly more likely to be causal than non-actions, $t(28) = 2.36$, $p < 0.03$.

Ratings from 37 participants in the *coherence condition* were analyzed (data from an additional three participants was discarded). As predicted, there was an overall significant effect of sequence type $F(2,70)= 9.18$, $MSE= 14.47$, $p < 0.0003$, with actions rated as going together significantly better than part-actions or non-actions $t(36)= 3.87$, $p < 0.0005$, one sample t-test on contrast values. There was also a marginally significant difference between part-action and non-action ratings $t(36)= 2.0$, $p = 0.05$.

Discussion

The results of this experiment support the hypothesis that people experience sequences of action grouped only by their statistical regularities as casually significant, meaningful groupings. Participants rated actions as more likely to cause a hidden musical effect than part-action and non-action sequences, even though all sequences were equally arbitrary, and in fact the non-actions for one exposure corpus were the actions for the other, meaning that the same sequences reversed their rating merely based on the number of times the SMEs appeared together. Similarly, participants rated actions as going together (a question we used as a measure of sequence coherence and meaningfulness) significantly better than other sequences. Anecdotally, a number of participants reported a feeling that the action sequences made more intuitive sense to them than the other sequences. Finally, all three conditions replicated the finding by Baldwin et al. (2008) that adults are able to parse statistically grouped actions from within a longer action sequence, and differentiate them from other non-action groupings, and confirmed the use of ratings as a viable alternative measure to forced choice comparisons.

These results have several important implications. First, they demonstrate that people's sensitivity to the statistical patterns in the exposure corpus is not simply an artifact of the impoverished stimuli, but appears to play a real role in their subsequent understanding of the intentional structure of the action sequence. The fact that participants found the statistically grouped actions to be more coherent, suggests that they do not experience the sequences they segment out as arbitrary, but assume that they are meaningful groupings that play some (possibly intentional) role. This is further supported by the results from the causal condition which show that, even without being presented with overt causal structure, people believe the statistically grouped actions are more likely to lead to external effects in the world.

Finally, these results also support our hypothesis that inference of action structure and causal structure are linked, with statistically grouped actions being perceived as more likely to also be causal variables. This result is consistent with our computational model, which also predicts that, without other evidence of causal structure, actions are more likely to be causal than non-action and part-action sequences.

Experiment 2: Using Causal Structure

Our second experiment investigated whether people are able to pick out causal subsequences from within a longer stream of actions, and whether they use this causal information to inform their action segmentations. Specifically, we hypothesized that when statistical cues to action segmentation are unavailable, adults will be able to use causal event structure to identify meaningful units of action.

Method

Stimuli The structure and stimuli for this experiment closely matched that of Experiment 1. However, in Experiment 2, there were no *a priori*, statistically-grounded actions. Instead, the exposure corpus was assembled using four SMEs, so that each individual SME would be seen an equal number of times, and all possible length three sequences of SMEs would also occur with equal frequency (see Figure 4). Throughout the exposure corpus, no length three subsequences containing repeats of an SME were allowed to occur. This resulted in 24 possible SME triplets. A target triplet of SMEs was then randomly chosen as the "cause". Whenever this sequence of motions was performed in the exposure cor-

Figure 4: A portion of the Experiment 2 exposure corpus. four SMEs (Poke, Look, Feel, Rattle) are distributed so that all possible triplets appear equally often. A target triplet (Look, Feel Poke) is chosen to cause a sound.

pus, it was followed by the object playing music (participants were able to hear the music, unlike in Experiment 1).

The exposure corpus was created by first generating 24 shorter video clips. Each clip was designed to have a uniform distribution of both individual SMEs and of SME triplets. Specifically, in each clip, the four unique SMEs appear exactly six times each, and 23 of the 24 possible SME triplets appear exactly once each. These 24 video clips were shown consecutively in the exposure corpus, but were clearly separated from each other by text notifying the participant of the beginning and end of each shorter clip. The result was an exposure corpus composed of 24 short video clips, with each SME appearing 576 times throughout the complete corpus, and each triplet appearing 20 to 24 times.

iMovie HD was used to assemble the exposure corpus and add a cartoon sound effect following every appearance of the target sequence. Two different exposure corpora, each using a distinct set of four SMEs were created. Look, Poke, Feel and Rattle were used to create the first exposure corpus, with Look-Feel-Poke being the target triplet, and Read, Slide, Blow, and Empty were used to create the second exposure corpus, with Slide-Blow-Empty being the target triplet.

Participants and Procedure Participants were 100 U.C Berkeley undergraduates. Participants were divided into the same three conditions as in Experiment 1, with the difference that after viewing the exposure corpus, they rated all 24 possible SME triplets, and that all participants were told that certain action sequences caused the bottle to play music.

Results

We analyzed all results using 2×2 ANOVAs on exposure corpus (1 or 2) and sequence type (target, other). No effects of exposure corpus were found.

Ratings from 28 participants in the *familiarity condition* were analyzed (data from an additional two participants who rated all sequences identically as either a 1 or 7 was discarded). Contrary to our predictions, and the predictions of previous work, there was no effect of sequence type $F(1,26)= 1.58$, $MSE= 1.74$, $p > 0.22$. Participants rated the target sequence and the other SME triplets as equally familiar.

Ratings from 30 participants in the *causal condition* were analyzed. As predicted, there was a significant effect of sequence type $F(1,28)= 193.97$, $MSE= 310.439$, $p < 0.0001$, with the target sequence being rated as much more likely to lead to a musical sound than the other SME triplets.

Ratings from 35 participants in the *coherence condition* were analyzed (data from five additional participants was discarded). As predicted, there was a significant effect of

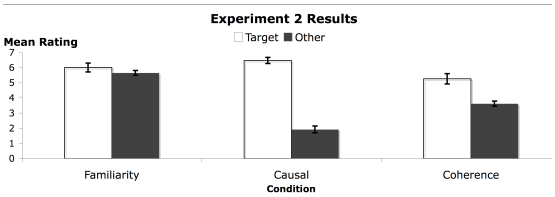

Figure 5: Results of Experiment 2. Error bars show one standard error.

sequence type $F(1,33)= 19.44$, $MSE= 47.1$, $p < 0.0001$, with the target sequence rated as going together significantly better than the other SME triplets.

Discussion

This experiment is one of the first to demonstrate that people can infer a correctly ordered set of causal variables from within a longer temporal sequence. In fact, the results of this experiment suggest that it was a relatively easy task for participants. Participants in the causal condition were nearly at ceiling in their ratings of how likely sequences were to lead to a musical effect, with the target sequence having a mean rating only slightly below 7 and the remaining sequences being rated a bit below 2.

The results of this experiment also provide further support for a relationship between action segmentation and causal inference. Even though there were no statistically grouped actions in this experiment, participants still perceived the target sequence as being more meaningful (going together better) than the other sequences, suggesting they had nonetheless segmented it out as a coherent action unit. It is worth noting that the ratings for the coherence question were different than those for the causal question, suggesting that participants did interpret the question as one of meaningfulness, rather than an alternate phrasing of the causality question.

Finally, it is interesting to note that, despite correctly identifying the target sequence as causal, participants *did not* rate it as more familiar than the other sequences. Instead, participants appeared to be aware that they had seen all the sequences an equal number of times, and rated them all as equally familiar. This implies that participants are not judging the target sequence as more coherent or more likely to be causal due to some sort of low level saliency effect that causes them to remember this particular sequence more clearly. It also suggests that participants, at least in this context, interpret the familiarity question as a question about frequency of appearance, which may help explain why Meyer and Baldwin (2008) failed to find sensitivity to conditional probabilities in action sequences using this question. These results suggest that participants may be aware that certain sequences are more causal or more coherent, while also being aware that they have seen other sequences equally often.

Modeling Segmentation and Causal Inference

We ran the model on the same exposure corpora our human participants watched in Experiments 1 and 2, to see if it could come up with the correct segmentation and causal

structure hypotheses. An abstract representation of each exposure corpus was used, with a letter standing for each SME. For each experiment, we ran two randomly seeded Gibbs samplers on each corpus, for 20,000 iterations. We then averaged results from 10 samples drawn from the last 1,000 iterations of each sampler, to estimate the posterior distributions and evaluate the model. For each experiment, results from both exposure corpora were combined. In addition, we evaluated the model's predictions for the coherence and causal conditions of our experiments, by representing the coherence of a sequence as its posterior probability of being in the inferred set of actions (referred to below as a lexicon), and its probability of causing an effect as the posterior probability of that sequence being followed by an event.

Experiment 1

We compared our results to the correct segmentation, and calculated average precision and recall scores across samples (commonly used metrics in the natural language processing literature). Precision (P) is the percent of all actions in the produced segmentation that are correct, while recall (R) is the percent of all actions in the true segmentation that were found. These scores are for complete actions, meaning that for an action to count, both boundaries must be correct. We also calculated average precision (BP) and recall (BR) for boundaries. Finally, we calculated precision (LP) and recall (LR) for the inferred action lexicon. As the results in Table 1 show, the model performed extremely well on all these measures of segmentation, especially when compared to a matched set of random segmentations.

Like our human participants, the model also predicts that actions are more likely to be in the lexicon, and more likely to be causal than non-actions and part-actions. However, the model's predictions are more extreme than human responses, with actions having an average probability of 0.88 of being in the the lexicon, and non-actions and part-actions never appearing at all. Similarly, the model predicts that actions are 5000 times more likely to be followed by an effect than part-action and non-action sequences. There are a number of possible reasons for this discrepancy. For instance, in addition to discovering complete actions, people may also be learning which motions are likely to appear together within novel actions. This would be equivalent to a model that learns the base distribution P_0, and could be explored in future work.

Model	P	R	BP	BR	LP	LR
Bayesian	0.83	0.75	1.0	0.88	0.75	0.88
Random	0.05	0.05	0.34	0.33	0.03	0.98

Table 1: Segmentation accuracy for Experiment 1 corpora.

Experiment 2

For Experiment 2, we were interested in seeing whether the model could infer the correct causal subsequence from within the longer sequence of motions (since the remainder of the input was noise, overall segmentation performance cannot be measured for this experiment). On average across samples, the model correctly segmented 76% of the occurances of the target sequence, while an equivalent set of random segmentations found only 7%. The model also correctly predicts that the target triplet is more likely to be in the lexicon than other triplets, with p(causal | lexicon)= 1.0 and p(other | lexicon)= 0.31, and is significantly more likely to be causal, with p(effect | target)= 0.999 and p(effect | other)= 0.0001. This performance is qualitatively very similar to that of our human participants in the Coherence and Causality conditions of Experiment 2.

Conclusion

People are able to use both statistical regularities and causal structure to help segment a continuous stream of observed behavior into individual actions. They can also identify the correct causal subsequence from within a longer set of motions. We used a non-parametric Bayesian model, adapted from work on statistical language processing, to infer the segmentation and causal structure of the same sequences our human participants saw. The parallels in both human and computational model performance between word segmentation and action segmentation tasks supports the possibility of a more general statistical learning ability. Future work will look at causal inference and action segmentation performance when the reliability of both sources of information is varied, and will explore the extent to which the model matches or differs from human behavior in more detail.

Acknowledgments. We thank Meredith Meyer, Kimmy Yung, James Bach, Mia Krstic and Jonathan Lesser, as well as the McDonnell Foundation Causal Learning Initiative Grant and Grant FA9550-07-1-0351 from the Air Force Office of Scientific Research.

References

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

Baldwin, D., Baird, J., Saylor, M., & Clark, A. (2001). Infants parse dynamic human action. *Child Development*, *72*, 708-717.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, *1*, 209-230.

Gilks, W., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. Suffolk, UK: Chapman and Hall.

Goldwater, S., Griffiths, T. L., & Johnson, M. (2006). Contextual dependencies in unsupervised word segmentation. In *Proceedings of Coling/ACL 2006.*

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*(5), 178-186.

Goodman, N. D., Mansinghka, V. K., & Tenenbaum, J. B. (2007). Learning grounded causal models. *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society*.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1-31.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354-384.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*(2), 72-79.

Meyer, M., & Baldwin, D. (2008). The role of conditional and joint probabilities in segmentation of dynamic human action. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*.

Saylor, M. M., Baldwin, D. A., Baird, J. A., & LaBounty, J. (2007). Infants' on-line segmentation of dynamic human action. *Journal of Cognition andDevelopment*, *8*(1), 113-128.