Learning Time-Varying Categories

Daniel J. Navarro (daniel.navarro@adelaide.edu.au)
School of Psychology, University of Adelaide, SA 5005, Australia

Amy F. Perfors (amy.perfors@adelaide.edu.au)
School of Psychology, University of Adelaide, SA 5005, Australia

Abstract

Many kinds of objects and events in our world have a strong time-dependent quality. However, most theories about concepts and categories are either insensitive to variation over time, or treat it as a nuisance factor that produces irrational order effects during learning. In this paper, we present a category learning experiment that explores people's ability to learn categories whose structure is strongly time dependent. In light of the results, we suggest that order effects in categorization may in part reflect a sensitivity to non-stationary environments, and that understanding dynamically changing concepts are an important part of developing a full account of human categorization.

Keywords: categorization, change detection, concepts, dynamics, time dependence, order effects

"Nothing endures but change." – Heraclitus

Categorization in a Non-Stationary World

At no two moments in time are we presented with the "same" world. Objects move, plants and animals are born and die, friends come and go, the sun rises and sets, and so on. More abstractly, while some of the rules that describe our world (e.g., physical laws) are invariant in our everyday experience, others (e.g., legal rules) are not. Given some appropriate time scale, certain characteristics of an entity or class of entities can change; moreover, they may tend to change in *systematic* ways. The event category of DAILY TEMPERATURES, for instance, has a natural yearly period and a gradual rising trend over the last 100 years due to anthropogenic global warming, in addition to geographic variation. In the context of familiar, everyday categories, people are highly sensitive to changes of this kind: if asked to predict the temperature 6 months from today, people will give quite different answers than if asked to predict the temperature tomorrow. That is, people do not simply modify predictions in an *ad hoc* or senseless fashion as the time of the future point draws ever more distant, as we can tell by comparing their predictions of the temperature in 12 months to the others. Rather, they appear to be attuned to particular details of the nature of the dynamic variation in category structure.

There are at least three ways that dynamic qualities might emerge as categories change over time. First, the characteristics of the individual entities that make up the category could each change over time. The social category of MY FAMILY has this property, for instance: even in the unlikely event that the membership does not change (no births, deaths or marriages), family members themselves grow and change over time. A second possibility is that the set of entities indexed by the category label can change over time, as when new members are added to a family. Another example of this is the natural category of PLANETS: in 2006 Pluto was officially removed from the category, after having been originally added in 1930. The third option is that the characteristics of items in the category can change due to some combination of the two: for instance, selection effects result in GIRAFFE necks becoming longer, or MOTH wings getting darker.

In addition, categories may differ in the *form* of their variation over time. For instance, many dynamic categories capture cyclical or sinusoidal variation: MONTHS, DAYS, and HOURS are all defined in terms of *where* in the cycle they occur as well as certain characteristic features. Sundays are defined as coming after Saturday and before Monday, and may contain features like "don't have to work", "go to church", or "have brunch with friends." Other categories might capture other sorts of variation. For instance, the category of CARS has seen a more-or-less steady change in some of the crucial features (e.g., "maximum speed", "quietness of engine", etc). Finally, in some categories the form of the variation may *itself* change over time. The category COMPUTERS shifted dramatically about 50 years ago, when the set of things indexed by the label jumped in a fairly *discrete* fashion from "people who calculate things" to "machines that calculate things". Since then, the feature values for digital computers have changed both in discrete ways (e.g., vacuum tubes were replaced by transistors) and continuous ways (the number of transistors has grown exponentially).

The Importance of Order

If the world has this dynamic quality – that is, if the observable structure of our experiences changes over time – then one of the major consequences for human learning is that the *order* of our observations matters. If told that the average temperatures over recent weeks week were 21, 25, 27, 30, 29, 33 and 32 (but did not know whether the scale was Celsius, Fahrenheit, or something else), the rising sequence makes it most likely that the season is SPRING; if told the same temperatures in reverse order, the most likely season would be AUTUMN. Accordingly, a sensitivity to the "dynamic" character of categories is of considerable value to any system that seeks to reason sensibly about a changeable world.

Despite its ubiquity and utility, dynamic variation in category structure is not typically taken into account in explanations or models of categorization. Order effects in categorization are themselves well-studied, but are generally viewed as

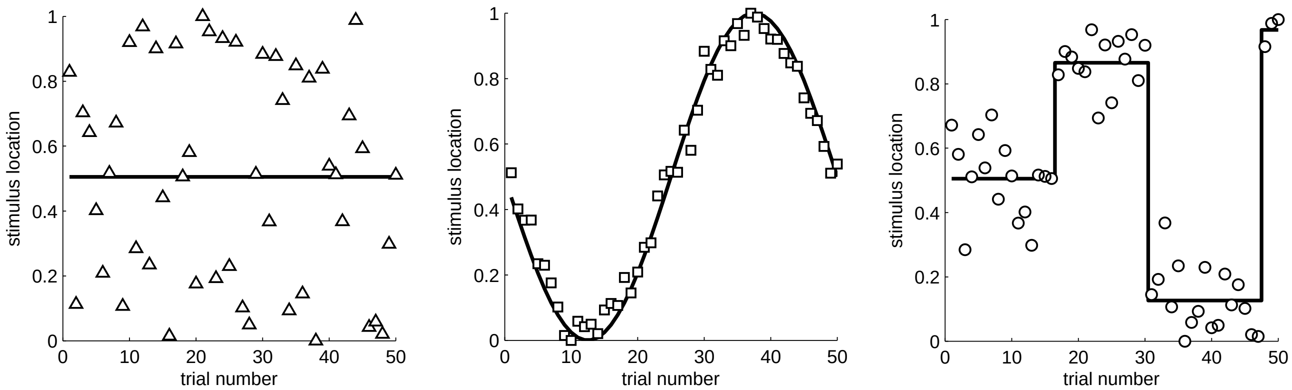

Figure 1: Data from the three categories used in the experiment (INDEPENDENT (left), SINUSOIDAL (middle) and DISCRETE JUMP (right)). All three consist of the same set of stimuli organized according to different kinds of sequential structure. The convention of using triangles to depict the INDEPENDENT category, squares for the SINUSOIDAL category and circles for the DISCRETE JUMP category will be maintained throughout the paper.

resulting from imperfections in memory and learning (e.g., Kruschke, 2006; Sakamoto, Jones, & Love, 2008). Whether these process limitations are seen to emerge due to the use of *ad hoc* (Anderson, 1990) or rationally motivated (Sanborn, Griffiths, & Navarro, 2006) computation strategies, it is implicitly assumed that in most cases people should *not* be sensitive to order information when learning new categories. While this is undoubtedly true in many cases, and we imagine that in general processing limitations play an important role during learning, it need not be universally the case. In fact, there are a number of cases in which these "limitations" might actually be sensible adaptations: for instance, forgetting old information is a reasonable strategy in a changing world (Anderson & Schooler, 1991), as is deliberately downgrading the value of such information (Welsh & Navarro, 2007).

As this discussion illustrates, one of the central assumptions in most descriptions of order effects is that they emerge because of the nature of the cognitive mechanisms or goals of the learner, rather than primarily due to the dynamic structure of the categories in the world. That is, in a categorization context, order effects are assumed to be arbitrary. In contrast, some recent research has suggested that the temporal structure of observations is crucial for rational learning: loosely mirroring ideas from the memory literature (Anderson & Schooler, 1991), when training data are autocorrelated in some fashion, then order effects are a hallmark of good reasoning, not bad (Yu & Cohen, 2009). However, even this does not capture the important insight that categories differ in the *form* of that autocorrelation, and that a reasonable learner should be sensitive to those dynamics as well.

In this paper we present data from an experiment in which people are presented with unidimensional stimuli that vary in particular time-sensitive ways. We show that people are, indeed, sensitive to this dynamic variation in category structure: in some instances the sequential structure leads people to (correctly) believe that the environment is highly predictable, while in other cases the structure can (again, correctly) lead people to suspect that future observations will be unrelated to the past. These results suggest that a full understanding of human categorization will require an understanding of how people think about dynamic as well as static categories.

Experiment

Our experiment is loosely inspired by the approach taken by Sakamoto et al. (2008), in which simple unidimensional categories are used, and the various category distributions differ only in terms of the order in which people observe the stimuli. We extend the design by (1) allowing for a broader range of sequential dependencies, (2) constraining the categories so that the sequential dependencies become necessary to differentiate the categories, and (3) using a predict-the-next task as well as a classification task. The rationale for incorporating the prediction task is to see if people are not just sensitive to sequential dependencies, but also able to extrapolate the underlying trends to the future. In short, we seek to discover the extent to which people can uncover and exploit category-dependent variations in their observations about the world.

Method

Participants. Thirty-two people were recruited from a paid participant pool largely consisting of undergraduate psychology students and their acquaintences. The experiment took place as part of a series of three unrelated studies, which took approximately 1 hour to complete. Participants were paid $12 for their time.

Category structures. Stimuli consisted of lines of different lengths presented on a computer screen; lengths varied from approximately 1cm (stimulus location "0") to 5cm (stimulus location "1").[1] All categories made use of the ambiguous distribution of category locations shown in Figure 2, but with three different orderings of stimuli. (That is, in all categories, the locations of the items were identical; categories differed only in terms of when during the presentation each item was shown). In the INDEPENDENT category, there was no time-dependent structure: the stimuli were ordered randomly. In the DISCRETE JUMP category, items from the middle of the location distribution were shown first, followed by items toward the upper end, and then items from the lower end, with the final three items being chosen from the top end. Finally, in

[1]Note that for half of the participants the mapping was reversed: stimulus location "0" corresponded to the longer lines, and location "5" to the shorter lines.

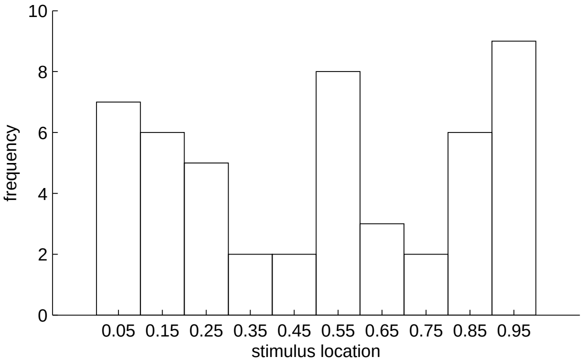

Figure 2: The marginal distribution of the locations of the category members is a noisy arcsin variate with additional mass near 0.5. The intent when constructing this distribution was that it be in itself somewhat ambiguous, and easy to convert to the three categories used in this experiment and shown in Figure 1.

the SINUSOIDAL category, the line lengths changed smoothly according to a sinusoidal function. The three categories are shown most clearly in Figure 1, which shows the data presentation as a function of time for each of them.

General procedure. Participants were randomly assigned either to a categorization condition or to a prediction condition. In both conditions, the cover story was constructed to allow for time-varying categories without explicitly drawing attention to the non-random ordering of items. The line lengths were tied to a pseudo-artifact cover story (a computer game) that suggested the existence of systematic rule-governed categories.

Categorization condition. The training phase for the categorization condition was a standard supervised learning task. The instructions in this condition were:

> Imagine that you're helping with the alpha testing for a new iPhone game. When finished, the game is going to involve things called WUGS and things called DAXES, and players of the game will need to learn which is which. At the moment, the developers don't have any flashy graphics, but they are testing some ideas about how DAXES and WUGS differ. So, for the moment, they're trying to figure out how hard or how easy different "DAX–WUG rules" are. With that in mind, they've put together a demo in which DAXES and WUGS are just lines on the screen, and they'd like you to try to figure out which is which, using the length of the line as a cue.

The onscreen display was designed to mimic the appearance of a mobile phone. Participants were shown a line and asked to guess the label. They responded using the keyboard, and received immediate feedback as to the correct label. Half of the lines belonged to the SINUSOIDAL category, and half to the DISCRETE JUMP category. These items were randomly interleaved: the complete sequence of 100 items is shown in the left panel of Figure 3.

After the training phase was complete, participants were asked to classify an additional 15 transfer items as DAXES or WUGS, and in this case no feedback was given. The transfer items were presented in a random order, and covered most of the range of possible line lengths in the task (though due to a coding error the transfer items were slightly "off-center"; see right panel of Figure 3). Before these were presented, however, participants were explicitly told that the "programmers" in the cover story had no clear intention about what should come next, and were primarily interested in soliciting opinions rather testing any explicit idea about what the "right" answer should be.

Prediction condition. In the prediction condition, participants were shown the stimuli in all three categories (i.e., including the INDEPENDENT category as well as the SINUSOIDAL and DISCRETE JUMP categories). On every trial they were shown a line and its accompanying label (either DAX, WUG or FAF) and asked to predict the length of the next line, which would be a member of the same category.

Instructions in this condition were thus similar to the instructions in the categorization conditions, except that the opening scenario involved FAFS as well as WUGS and DAXES. Also, instead of asking people to make classification decisions, the stimuli were labelled, and participants were asked to predict the length of the next observation of each. Specifically, they were told that they would

> be shown a coloured "WUG" line on the screen, and you'll be asked to guess how long the next WUG will be, which you can do by positioning the crosshairs on screen and clicking the mouse. You'll see a series of 50 DAXES, followed by 50 WUGS and then 50 FAFS, so in total you'll need to make 150 decisions.

After being shown all 50 items in each series, participants were asked to predict the lengths of the next 5 members, but were not given feedback.

Results

We consider the categorization data first, which present an odd puzzle, and then turn to the prediction data, which help to resolve it.

Categorization condition. Figures 3 and 4 shows the general pattern of results for the categorization condition. The plot on the left hand side of the Figure 3 shows a condensed description of the training data in which white-colored markers denote trials in which people performed better than chance, and black markers display trials in which performance was below chance (trials that were indistinguishable from chance are not shown). Figure 4 expands this somewhat, plotting the average probability of a correct response for every trial in the experiment.

To determine which trials were at chance, which were above and which were below, we used a simple Bayesian data analysis method involving three hypotheses about θ_t (the probability of a correct response on trial t). The chance hypothesis is $H_0 : \theta = \frac{1}{2}$, while the two non-chance hypotheses are $H_+ : \frac{1}{2} < \theta \leq 1$ and $H_- : 0 \leq \theta < \frac{1}{2}$. For the two non-chance hypotheses, we assume a uniform prior over the admissable values of θ (which makes the model a incomplete beta-binomial model, and straightforward to evaluate; see, e.g., Gelman, Carlin, Stern, and Rubin (1995)). We assume that each hypothesis is equally likely *a priori*, and choose the one that is most likely having observed the data. It is this analysis that produces the colorings shown in Figure 3.

The central point is that the sequential dependencies are clearly strong enough for the distinct categories to be distinguishable from each other, even though they both consist of the exact same set of entities. This is in part because on

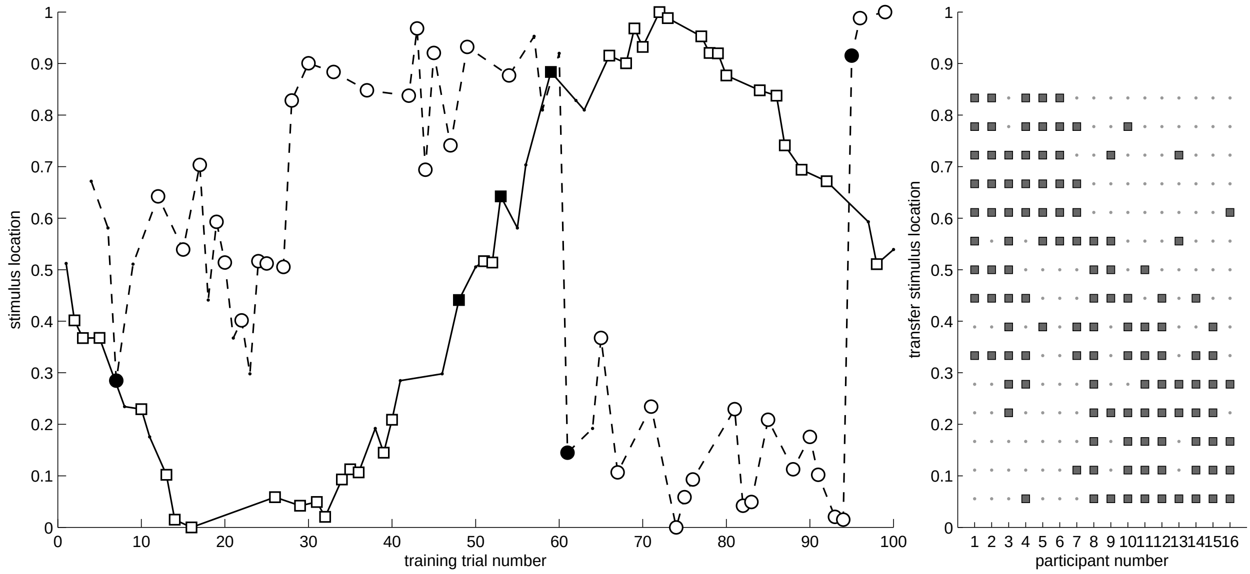

Figure 3: Categorization condition results. Data from the training phase (left) and transfer phase (right). For the training data, circles denote items belonging to the DISCRETE JUMP category, and squares show items belonging to the SINUSOIDAL category. The white-colored markers correspond to trials in which participants' classification decisions were better than chance, whereas the black-colored markers display those trials where people performed below chance. On some trials performance was statistically indistinguishable from chance levels: no markers are plotted for those trials. Despite the fact that both categories index the exact same collection of objects (Figure 2) and are differentiated only by the time-dependent order effects, participants generally perform well. For the transfer data (right panel), the grey squares denote stimuli that people classified as belonging to the SINUSOIDAL category, with dots marking the other trials.

any given trial the conditional distribution over the current observation for the two categories is negatively correlated ($r = -.47$), which provides some basis for distinguishing between the two. However, in order for people to exploit this correlation, they need to be able to predict correctly *where* at least one of the categories is currently generating data – otherwise the correlation is useless. The sequential dependencies are critical for this purpose, and people are clearly able to exploit them, as illustrated on the left panel of Figure 3. That is, the fact that most markers are white-colored implies that on most trials people possessed some knowledge about the category label.

Despite the evidence that participants appear to exploit order effects during learning, the transfer data appear on first glance to suggest that they fail to do so during transfer. The columns in the right panel of Figure 3 show the transfer classifications of each participant. As is evident, most participants produce internally consistent transfer data in which shorter lines are assumed to belong to one category and longer lines to the other – but there is no consensus *between* participants as to which is which.

These results present us with something of an oddity. On the one hand, people must be able to uncover and use the sequential dependencies, since they are clearly able to learn the categorization rules during the training phase.[2] However,

[2]Note that the data do not determine whether people learn that each category changes over time, or merely that or merely that the learned rule about DAXES and WUGS flips. Either way, people are sensitive to time-dependent variation, so we leave the issue of the precise *nature* of this sensitivity for future work.

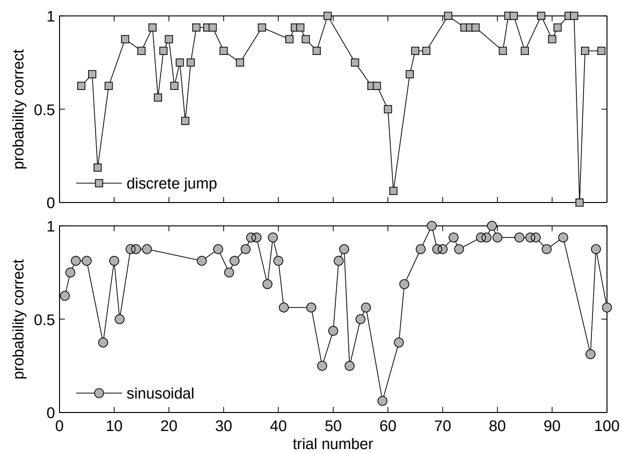

Figure 4: Probability of a correct classification, as a function of trial number and category. Chance is 50%.

whatever learning has taken place does not seem to lead to any consistent pattern of discrimination between the categories on transfer. To resolve this anomaly, we turn to the data from the prediction condition.

Prediction condition. Figures 5-7 shows the average predictions made by people during the training phase (left panels) and their typical predictions in the transfer trials (right panels). In each figure, the solid line in the left panel indicates participants' predictions at each point; the predictions made on the 5 transfer trials are summed up in the histogram in the right panel. For instance, the right panel of Figure 5

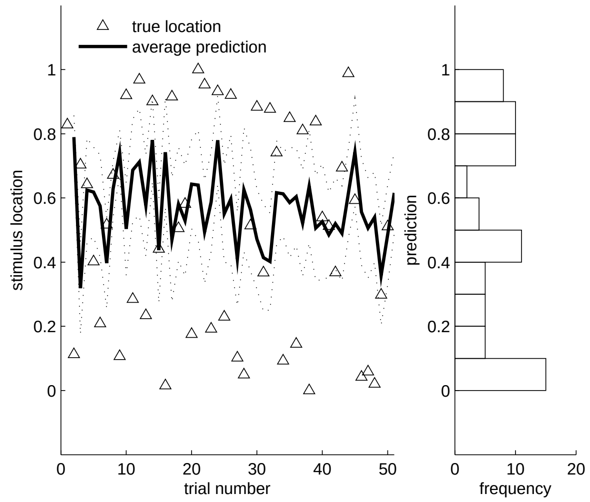

Figure 5: Prediction condition: INDEPENDENT category. The left panel shows the average sequential prediction strategy used by people (solid line) and corresponding 95% confidence intervals (dotted lines), plotted against the true locations (circles). The right panel shows the distribution over predictions made on the transfer trials. Not surprisingly, the average prediction on a trial-to-trial basis shows no pattern. What matters, however, is that the transfer trials fairly closely reproduce the marginal distribution in Figure 2.

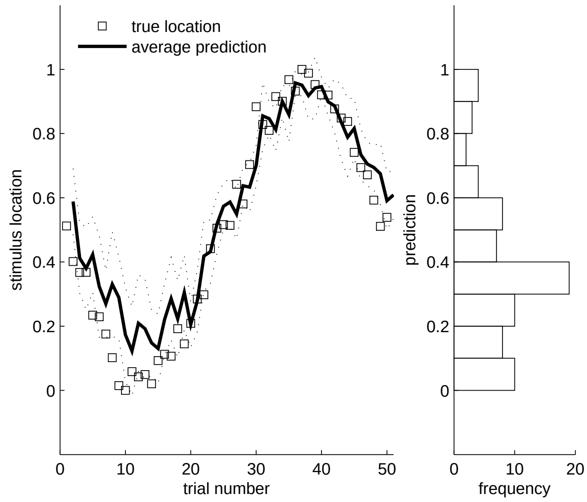

Figure 6: Prediction condition: SINUSOIDAL category. The left panel shows the average sequential prediction strategy used by people during training, and the right panel shows the corresponding transfer generalizations. People accurately track the sinusoidal variation as one might expect, but more importantly the distribution on transfer has a genuine *predictive* quality, since the typical transfer location prediction is lower than the location of items in the most recent trials.

shows the modal prediction to be between location 0 and 0.1.

A comparison of Figures 5-7 provides a robust indication that participants are successfully categorizing on the basis of the time-dependent presentation of items (if such dependency exists). Results from the INDEPENDENT category, shown in Figure 5, demonstrate that when there is no time-dependent category structure, participants show no pattern to their predictions, whether during training or transfer. Indeed, the distribution of predictions about item location during the transfer trials closely matches the distribution of item locations during training (as shown in Figure 2): participants are not inferring any additional pattern.

By contrast, results from the SINUSOIDAL and DISCRETE JUMP categories indicate that participants were sensitive to the distinct time-dependent category structure of each. Figure 6 illustrates that people clearly understood the sinusoidal structure of the category during training, and their performance on the transfer trials demonstrates that they are using this structure to correctly predict what they would see next. The transfer performance is especially interesting because simple heuristics like "predict what has been happening" would not capture what humans are doing here, since they (correctly) extrapolate that the next items should be found at a location lower than any of the most recent ones.

Figure 7 is interesting because it demonstrates an apparent divergence between training and transfer performance (and, thus, an explanation of participant behavior in the categorization condition). The training data indicates that participants were able to induce the time-dependent structure of the category reasonably well, although they showed considerable uncertainty about the sudden shift occurring at the very end of the sequence. This is sensible, because there are only a few trials' worth of data after that shift, making it unclear as to whether those datapoints indicate a "real" shift (like the one that occurred around trial 30) or not. This uncertainty is evident in the transfer data, which show a high degree of entropy. The transfer predictions do not match the original location distributions (as in Figure 2), suggesting that participants know there is *some* time-dependent structure, but also do not reflect coherent beliefs about the future (as in the SINUSOIDAL category shown in Figure 6).

This may explain performance in the categorization condition, where we observed that most participants produced internally consistent data and tended to assume that shorter lines could be classified into one category and longer lines into another. As Figure 6 makes clear, participants learned a highly consistent predictive model for future data generated from the SINUSOIDAL category, but did not appear to do so for the DISCRETE JUMP category (Figure 7). Presumably, the fact that the future behavior of the category was well-understood by people only in one case made the transfer task in the categorization condition quite difficult.

General Discussion

These results demonstrate that human learners are quite sensitive to time-dependent variation in category structure, and we suggest that this sensitivity is not always a result of characteristics of memory and learning, such as processing limitations or rational discounting of past information. Rather, because the observable structure of our experiences changes over time, a rational learner should be attuned to that variation and be able to use it where it is relevant. Our experiment offers a demonstration that at least in this very simple case, humans are surprisingly successful at doing this.

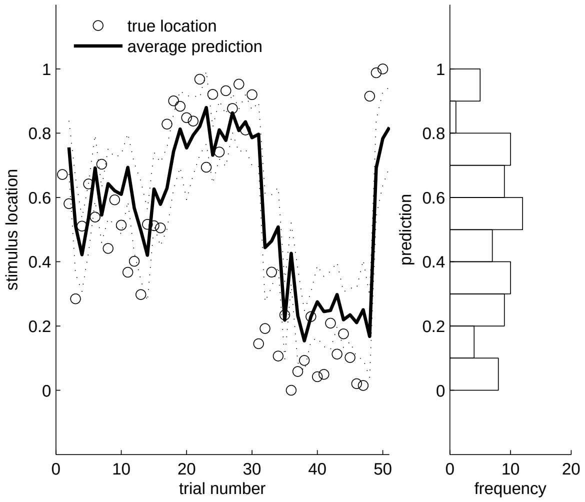

Figure 7: Prediction condition: DISCRETE JUMP category. The left panel shows the average sequential prediction strategy used by people during training, and the right panel shows the corresponding transfer predictions. The predictions in this case are reasonable, though it is clear that there is considerable uncertainty about the sudden shift that occurs at the end of the sequence: the average prediction at the end is regressed a long way to the middle. This uncertainty is reflected in the transfer predictions, which do not reflect either the marginal distribution (as per Figure 5) or any coherent belief about the future (as per Figure 6).

This work opens a broad avenue of future directions. On the experimental side, it is important to follow up this work in situations involving richer categories. Are people so quick to induce time-dependent structure when there are other important features as well? For instance, if instead of being shown lines differing along only one dimension (location), what if people were shown items differing along many features (color, shape, texture, and location), only one or a few of which varied consistently over time? Would it make a difference if the time-dependent variation occurred over a short scale (in which case it might be automatically detected by the low-level visual system) or over a very long scale (in which case memory limitations might apply)?

On the theoretical side, this work suggests that a complete model of human categorization should include a component that can account for people's sensitivity to dynamic structure. We presume that this could be added to many current approaches of categorization (see Kruschke, 2008), and suggest that work along these lines could be further used to distinguish the advantages and disadvantages of each type of model. For instance, the models used by Sanborn et al. (2006) and Sakamoto et al. (2008) can both be characterized as methods for "tracking" an estimate of a category distribution. In the original models, the category itself is not assumed to change, only one's knowledge of it. However, as discussed by Arulampalam, Maskell, and Gordon (2002), it is not difficult in principle to extend these approaches to a "predictive tracking" model, in which the learner allows for the world to change over time (see, e.g., Freyd & Jones, 1994).

More broadly, our work moves a step beyond assuming that categorization consists only of noticing regularities in observable features. The idea that categorization can also occur on the basis of regularities over *time* may provide a way to synthesize areas in cognitive science that are typically seen as distinct. For instance, the study of linguistic knowledge and use is focused on understanding how humans categorize a particular sort of time-dependent variation (namely, sequences of words or phonemes). Regardless of whether the same sorts of cognitive abilities that underlie categorization of non-linguistic time-dependent regularities also apply to linguistic ones the answer promises to add a great deal to our understanding of language as well as categorization.

Conclusion

In sum, these results show that human learners are capable of learning time-dependent category structure. We suggest that a rational learner should be sensitive to such structure, since sequential structure is an essential characteristic of both natural categories (e.g., SPRING and AUTUMN) and created categories (e.g., BULL MARKETS and BEAR MARKETS). Moreover people are appropriately influenced by the *form* of the dependency – assuming that COMPUTERS change like SEASONS would be inappropriate. As a consequence of this sensitivity, we suspect that order effects in categorization may not always be entirely due to processing or memory limitations. As we move toward a fuller understanding of human categorization, people's sensitivity to this sort of information needs to be explained.

Acknowledgements. We thank Carolyn Chen, Steven Langford and Xin Wei Sim for helping to run the experiments, and reviewers for their helpful comments. DJN was supported by an Australian Research Fellowship (ARC grant DP0773794). Participant payments were partially supported by the Centre for Human Decision Making and Expertise at the University of Adelaide.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396-408.

Arulampalam, S., Maskell, S., & Gordon, N. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*, 174–188.

Freyd, J. J., & Jones, K. T. (1994). Representational momentum for a spiral path. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 968-976.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.

Kruschke, J. K. (2006). Locally Bayesian learning with applications to retrospective revaluation and highlighting. *Psychological Review*, *113*, 677-699.

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The cambridge handbook of computational psychology* (p. 267-301). New York: Cambridge University Press.

Sakamoto, Y., Jones, M., & Love, B. C. (2008). Putting the psychology back into psychological models: Mechanistic versus rational approaches. *Memory and Cognition*, *36*, 1057-1065.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2006). A more rational model of categorization. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 726–731). Mahwah, NJ: Lawrence Erlbaum.

Welsh, M. B., & Navarro, D. J. (2007). Seeing is believing: Priors, trust and base rate neglect. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 701–705). Austin, TX: Cognitive Science Society.

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, *21*.