Anticipating changes: Adaptation and extrapolation in category learning

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

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

Abstract

Our world is a dynamic one: many kinds of objects and events change markedly over time. Despite this, most theories about concepts and categories are either insensitive to time-based variation, or treat people's sensitivity to change as a result of process-level characteristics (like memory limits, captured by weighting more recent items more highly) that produce irrational order effects during learning. In this paper we use two experiments and nine computational models to explore how people learn in a changing environment. We find, first, that people adapt to change during a category learning task; and, second, that this adaptation stems not only from weighting more recent items more highly, but also from forming sensible anticipations about the nature of the change.

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

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 – like physical laws – are invariant over the course of our everyday experience, others – like 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. Temperature varies cyclically as a function of time of day and time of year; with a gradual rise over the last century. As another example, the features that describe phones have changed considerably over recent decades: not only do modern phones perform many new functions, they are also physically smaller, sleeker, and smoother. People are aware of this temporal structure and adapt their predictions to it: if asked to predict the temperature six months from today, people will give quite different answers than if asked to predict the temperature tomorrow. We do not modify predictions in an *ad hoc* or senseless fashion; instead, we appear to be attuned to particular details of the nature of the dynamic variation in category structure.

One consequence of recognizing the changeable nature of categories is that the time at which observations are made matters. A machine with a rotary dial could very well be a typical phone if it were observed in 1980, but this is much less plausible if the observation dates from 2010. These changes over time introduce strong *order effects* into the classification problem. Order effects in categorization are well-studied, but people's sensitivity to order is generally thought to result from weighting more recent items more highly (generally as a result of poor memory) or inefficient learning rules (e.g., Kruschke, 2006; Sakamoto, Jones, & Love, 2008). Irrespective of whether these process limitations arise from the use of *ad hoc* (Anderson, 1990) or rationally motivated (Sanborn, Griffiths, & Navarro, 2010) computation strategies, the shared assumption is that people should *not* be sensitive to order information when learning new categories. While this is certainly true in some cases, in other situations order sensitivity might actually be a sensible adaptation. Especially when a category is changing in systematic ways (e.g., phones are getting steadily more complex), a sensible learner should be able to extrapolate about the future in a way that is sensitive to that systematicity. The central question of this paper is whether people can perform this extrapolation, and whether it can be separated from an order sensitivity that arises from simply weighting more recent items more highly.

The structure of this paper is as follows. The first half explores the advantages of being sensitive to order in category learning. This is important because the literature tends to focus on the negative consequences of order sensitivity – namely, irrational order effects when the environment is in fact stationary. However, we present computational modeling highlighting the fact that when the world is actually changing, being sensitive to that change constitutes an enormous advantage. In the second part of the paper, we investigate the nature of human order sensitivity. To what extent does it reflect imperfect memory, and to what extent does it reflect sensible adaptation and extrapolation about systematic change? Experimental and computational results[1] demonstrate that both factors play an important role.

The importance of order sensitivity

We begin our exploration into the importance of order sensitivity with a prototype model for categorization, which in its standard form is not sensitive to order. In this model, people are assumed to construct a single representation of the prototype, or the central tendency of the category. In its most usual form, if the learner has seen a series of t items, then the prototype μ_t is found by taking the arithmetic mean of the mental representations of the category members:

$$\mu_t = \frac{1}{t}\sum_{i=1}^{t} x_i \quad (1)$$

where x_i denotes the mental representation (e.g., co-ordinates in psychological space) of the i-th category member. If the

[1] Because the paper involves a large number of models (9 in total), we outline the structure of the models only in general terms here: precise specifications are available in the supplementary materials found at `www.compcogscilab.com/dan/`.

distribution of the category members is normal, then the learner is assumed to keep track the standard deviation of the t category members around the prototype σ_t. The probability that a new item belongs to the category is proportional to the probability that the category distribution assigns to that item. As noted by several authors (e.g. Ashby & Alfonso-Reese, 1995) this model has an interpretation as a rational model when the category members satisfy certain assumptions – including the assumption that the world does not change.

It is precisely because the world is assumed not to change that all category members are assigned equal weight in Equation 1, regardless of when they were observed. As a consequence, the model predictions on trial t do not depend on the order in which the preceding items were observed. This order invariance is characteristic of most probabilistic models for category learning, whether they be prototype, exemplar, or cluster-based (see Griffiths, Sanborn, Canini, & Navarro, 2008). As noted by a number of authors (e.g. Sakamoto et al., 2008; Navarro & Perfors, 2009) such models perform poorly as models of human performance when the data presented to people contains strong order effects. To address this, it is commonplace to consider an updating rule that is more sensitive to the changing structure of the data. The general idea is to assume that when the learner encounters a new category member, he or she moves the prototype a little closer to the new category member. This has the effect of weighting new items more than older items. The simplest version of this rule is as follows: if x_t refers to the feature value(s) for the new item, then the new prototype location at time $t+1$ is

$$\mu_t = \phi\mu_{t-1} + (1-\phi)x_t. \tag{2}$$

One advantage of this rule is that it is simple and easy to implement. Viewed from a computational perspective, it corresponds to an "exponential weighting" scheme, in which the most recent observations are weighted most highly, where the weight decays exponentially as a function of time.[2] In that sense, it is quite similar to the exponential generalization gradients that are typically used in category learning models (Nosofsky, 1984), but applied across time rather than psychological space. Additionally, this weighting function is similar to the shape of the retention function in memory research.[3]

The key feature of this and similar rules is that the learner is constantly adapting the prototype, never converging to a single true value. Because of this, the model is quite sensitive to the recent trends in the data. Learning rules of this form capture human order effects in simple choice tasks (e.g., Yu & Cohen, 2009; Gökaydin, Ma-Wyatt, Navarro, & Perfors, 2011). Moreover, such rules make sense as a near-optimal inference scheme when there is a constant probability that "something changes", but the nature of the change is not predictable or systematic (e.g., Yu & Cohen, 2009; Brown & Steyvers, 2009). In the next section, we illustrate how important this sensitivity can be.

[2]The exact equation is $\mu_t = (1-e^{-\tau})\sum_{i=1}^{t} e^{-\tau(t-i)}x_i + e^{-\tau t}$, where $\tau = -\ln\phi$, and μ_0 is the initial location of the prototype. The derivation is included in the supplementary materials, but since Equation 2 is the simpler and more interpretable version, we use it throughout the paper rather than this more explicit exponential form.

[3]Although retention curves are generally have heavier tails than exponentials (e.g. Rubin, Hinton, & Wenzel, 1999), this would likely make little difference to a typical category learning experiment.

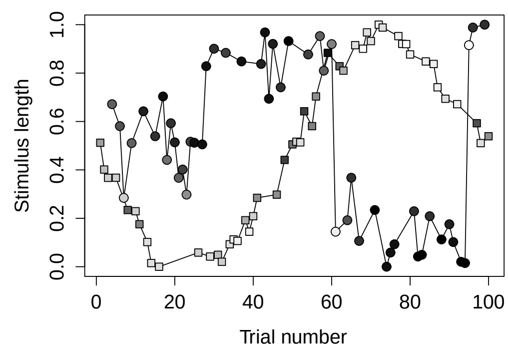

Figure 1: The experimental design from Navarro and Perfors (2009). Stimuli varied along a single dimension (line length). Circles denote items belonging to one category, and squares refer to the other category. The two categories consist of the exact same set of 100 stimuli, but are presented in a different order. The shading summarises the human responses: the darker the shade, the more likely people were to select the "circle" category (see Figure 2 for histograms showing the probability of a correct response). Note the large "jump" around trial 60, and the fact that humans quickly adapt to it.

Modelling adaptation to change

The typical view of "rationality" in a category learning experiment assumes that the world is stationary. Under this assumption, any sensitivity to the order in which items are observed is irrational. However, this view is sometimes expressed even when using experimental designs that *violate* the stationarity assumption. A good example of this is the paper by Sakamoto et al. (2008), which presented results from a simple supervised categorization task in which one category possesed a strong time dependency. In this task, an order sensitive model accounted for human classification decisions better than an order insensitive one. Despite the fact that is sensible to be sensitive to order such a situation, this was argued to be evidence *against* the rationality of human behavior rather than evidence in favor of it.

To explore the extent to which people are sensitive to categories that change over time, Navarro and Perfors (2009) conducted a categorization experiment in which the *only* information that distinguished the categories was the order in which items were presented (the design is illustrated in Figure 1). Human performance was well above chance: participants correctly classified the stimuli on 76% of trials. However, models were not fitted to this data, making it somewhat unclear exactly how badly an order insensitive model would perform on the task, and the extent to which order sensitive models would improve matters.

To address these question, here we fit[4] six different models to that data. Three of the models are not sensitive to order. This includes two variations of a standard prototype model, one assuming that both category distributions have equal variance (i.e., the same value of σ) and the other allowing unequal variances, and a standard exemplar model (Nosofsky, 1984). The other three models are analogous, but each use an exponential weighting scheme to assign more importance to recent

[4]Throughout the paper we use the ordinary least-squares (OLS) method to estimate model parameters: that is, we minimize the sum squared deviation between model predictions and human classifications, though we exclude the first 10 trials since human and model predictions are both quite variable initially.

(a)

(b)

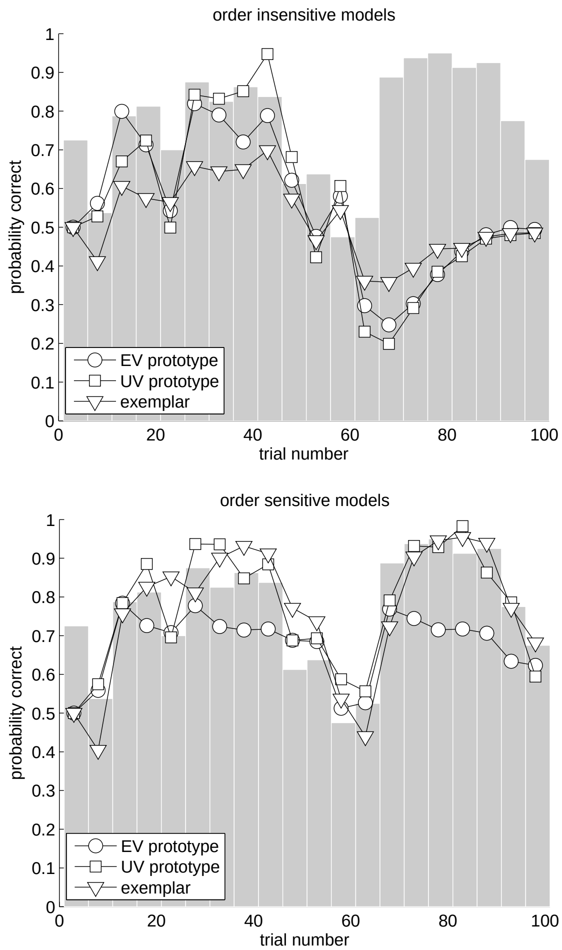

Figure 2: Graphical depiction of the performance of prototype (EV = equal variance, UV = unequal variance) and exemplar models on the classification task, for both the standard order-insensitive versions (panel a) and the order-sensitive versions (panel b). The grey bars plot human performance, and for reasons of visual clarity all plots are averaged over 5 trial blocks. Notice the catastrophic prediction failure for the order insensitive models during the second half of the experiment.

items: this is equivalent to Equation 2 for the prototype models, and is slightly more complex for the exemplar model. The formal details for all models are included in the supplementary materials. None of the models have more than two free parameters in this task.

The important finding here, shown in Figure 2a, is that none of the order insensitive models can mimic human behavior, nor can they achieve good classification performance on the data. No order insensitive model can correctly classify the stimuli at a rate higher than 55%, regardless of what parameter values are chosen. Moreover, none of the models can account for more than 5% of the variance in human responses.[5] The basic problem is evident in Figure 2a: when the task changes at trial 60, no order insensitive model can adapt to it in the way humans do. Viewed as *psychological* models, the inability to describe human behavior is a serious issue. In addition, the inability to achieve an acceptable classification performance makes it clear that none of these models provides a good *rational* standard for the task either.

In stark contrast, all three of the order sensitive models perform reasonably well, as shown in Figure 2b. When fit to the human responses, the exemplar model and unequal-variance prototype model are nearly indistinguishable, accounting for 79% and 80% of the variance respectively. The equal variance prototype model performs slightly worse, accounting for 72% of the variance. This is not surprising: as Figure 1 illustrates, the categories are not equally variable. In any case, the important point is that all three order sensitive models are massively superior to any of the order insensitive models.

This improvement in psychological performance is matched by an improved classification rate. If the parameters are chosen to maximize the classification accuracy rather than to maximize the variance accounted for in human performance, the exemplar model and unequal variance prototype models can both exceed human performance (89% and 82% accuracy), while the equal variance prototype model is slightly worse (68% compared to the 76% of humans). While it would be incorrect to suggest that any of these models provides an exact normative standard for this task, since the data have a rather more complex structure than a simple exponential weighting mechanism suggests, it is clear from the superior classification performance that the order sensitive models are much closer to the required standard than the usual order insensitive models.

We have shown so far that order sensitivity is important in understanding human category learning: standard order insensitive models fail catastrophically when the world changes, both as pure classifiers as well as models of human classification. By contrast, models that use a simple exponential weighting method do surprisingly well in capturing human order sensitivity. But is that the *entire* explanation for how humans respond to time dependence in the input? One possibility is that it is, and that people change over time because they have poor memories and are most likely to recall and use more recent information. The other possibility is that when the changes are systematic, people actively notice and adapt to them, allowing them to extrapolate in a sensible way when making predictions about new items. In the rest of this paper we describe experimental and computational work showing that both possibilities appear to be true: human performance is affected by memory limitations as well as the ability to adapt in a sensible way.

Learning to predict changes

The central question going forward is whether people's sensitivity to change is the result of memory limitations or to a sensible adaptation to a changing world. To some extent, these explanations are not in conflict: memory is sensitive to the need probability of the information to be recalled, and recent information is more likely to be useful than older information when the world changes (Anderson & Schooler, 1991). Nevertheless, it is possible to make distinctions between the two. For instance, when the world changes in *systematic* ways, it is not usually an optimal strategy to just reweight old informa-

[5] We measure this using the Variance Accounted For (VAF) statistic, which is equivalent to an R^2 statistic in regression, but does not estimate an intercept and slope: the model predictions are compared *directly* to human performance.

Table 1: An overview of the five models used in the experiment. All models are equal-variance prototype models, but differ in the manner in which the category prototype is estimated. The first column gives the label for the model, the second column indicates whether the model used exponential weighting or not, and the third column indicates what method the model used to anticipate what future changes would occur.

Model	Weighting	Extrapolation
STANDARD	constant	none
RECENCY	exponential	none
RECENCY+BIAS	exponential	constant
REGRESSION	constant	regression
RECENCY+REGRESSION	exponential	regression

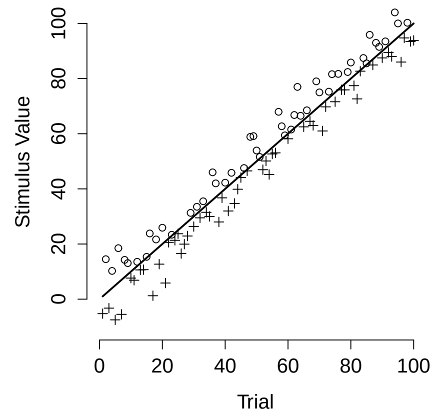

Figure 3: The experimental design. Circles denote stimuli belonging to the "high" category and crosses denote stimuli belonging to the "low" category. Although the classification rule changes over time – that is, the classification boundary is constantly rising – it does so in a regular fashion. The scale on the vertical axis is normalized so that the average "rise" from one trial to the next is 1 unit; on screen, this corresponded to an average rise of approximately 2mm per trial. Note that although it is logically possible to correctly classify all items, in practice the task is quite difficult, since most stimuli lie close to the boundary.

tion. Instead, the right strategy is to learn to anticipate those changes, and guess how the world is going to change before the change actually happens. In this section we present a category learning experiment in which people were presented with fairly obvious and systematic temporal changes, and investigate how well people learned to anticipate those changes.

Method

Participants 59 participants were recruited through a mailing list whose members consist primarily of current and former undergraduate psychology students, and paid $10/hour for their time. The median age was 23, and the participants were predominantly (63%) female.

Materials & Procedure The learning task was a standard supervised classification experiment, performed on a computer. Stimuli were little cartoon objects ("floaters"), which were displayed floating above a horizontal line ("the ground"). The height of the floater was the only respect in which the stimuli varied from each other.

On each of 100 trials participants were shown a single floater and asked to predict whether it would flash red or blue. After making their prediction, they would receive feedback for two seconds while the floater flashed the appropriate color. As Figure 3 illustrates, as the experiment progressed, all the stimuli shown to people tended to rise, regardless of which category they belonged to. In the figure, circles correspond to items that belong to the "high" category and crosses correspond to times that belong to the "low" category. Assignment of flash color (red or blue) to category (high or low) was randomized across participants.

The classification rule was such that, if x_t denotes the height of the stimulus on trial t, then the optimal response is to select the response option corresponding to the high category if $x_t > t$. Such a rule, shown as the solid line in Figure 3, achieves 100% accuracy on the task. However, because most stimuli tend to lie quite close to the classification boundary, the task is relatively difficult even though the general trend is clear. Consistent with this, participants during informal discussions indicated that they detected the upward trend early in the experiment, but still found the task to be quite challenging.

Models We fit five models, outlined in Table 1, to the experimental data. For this task a prototype model provides a good description of the category representation, and since the two categories are in fact equal variance we can use the equal variance prototype model as our basic model.[6] The only differences among the five models lies in how they calculate the location of the prototype μ_t; we systematically vary effects of memory (i.e., how previous items are weighted) as well as anticipation (i.e., what extrapolation method, if any, is used to anticipate future items).

The first two models, STANDARD and RECENCY, do no extrapolation at all: they are identical to the two equal-variance prototype models from in the previous section. The STANDARD model is the order-insensitive model with constant weights as in Equation 1 whereas the recency model captures order sensitivity via the exponential weighting scheme in Equation 2. A third model, which we call RECENCY+BIAS, incorporates this exponential weighting scheme, but also incorporates a component that captures participant adaptation to the experimental design. It does so by shifting the prototype upwards by a constant amount β. In other words, this model not only *follows* where the data are moving (using the exponential weighting scheme), but it *extrapolates* to guess where the data are moving to.[7]

The final two models explore the extent to which people use a different and more optimal extrapolation method. These models estimate an explicit linear regression model for each category prototype, using the trial number as the predictor (i.e., $\mu_t = at + b$), re-estimating the regression equation after each trial. This is not quite an ideal observer model for the task, but it is extremely close: if we were to allow it to respond deterministically (via the Luce (1963) choice rule, for

[6] We expect that the same pattern would be observed if we used exemplar models or unequal variance prototype models; nothing in the design is specific to the equal variance prototype model.

[7] Note that this is not quite identical to the usual manner in which bias parameters are used in category learning models. The bias in this case alters the underlying category representation (the prototype) rather than the response bias, although in this experimental design the effect is not dissimilar.

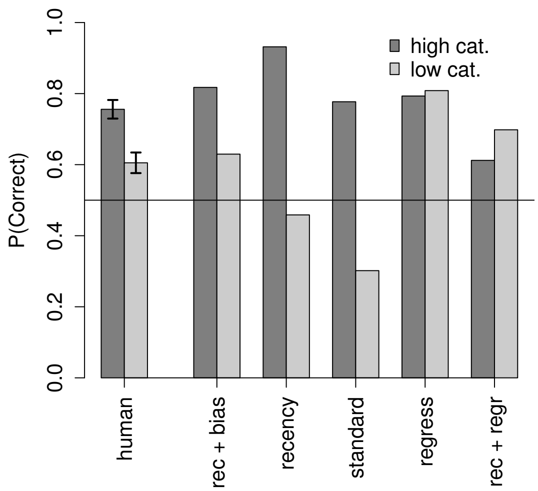

Figure 4: Overall performance for humans and models. The qualitatively important characteristics are that humans perform above chance on both categories, but perform better for the high category. This qualitative pattern is captured only by the RECENCY + BIAS model, which extrapolates about the future but also weights more recent items more highly.

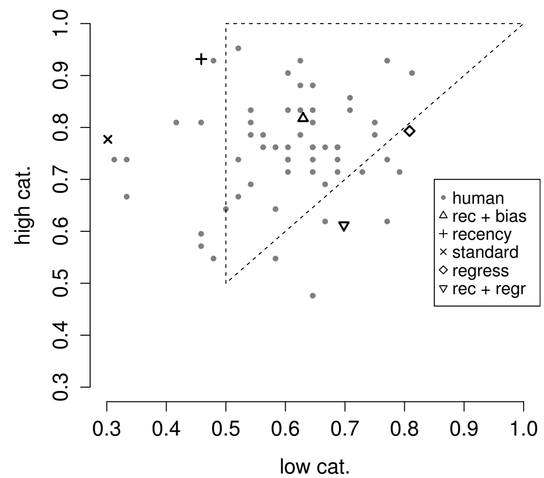

Figure 5: Individual participant data. Each dot represents one participant: the horizontal co-ordinate shows their classification performance on items that belonged to the lower category, and the vertical co-ordinate plots the performance on the higher category. The other markers show the model predictions, as indicated in the legend. The dotted triangular region corresponds to all possible patterns in which classification performance is above chance for both categories, but with the high category showing better performance. The majority of human participants fall within the triangle, as does the RECENCY+BIAS model.

instance), it would be able to perfectly classify all items. The fourth model, REGRESSION, uses exactly this linear regression model, while the fifth (RECENCY+REGRESSION) adds an exponential weighting scheme in order to attach more importance to recent observations when estimating the regression equation. In effect, this model assumes that the learner applies a nonparametric regression model (in this case lowess regression; Cleveland & Devlin, 1988) to estimate the location of the prototype, rather than a linear regression model.

The formal specification for all models can be found in the supplementary materials. For now it is sufficient to note that the only free parameters in the models are the ϕ parameter that governs the amount of weighting for the three models that use exponential weighting (Equation 2) and an additional bias parameter β for the RECENCY+BIAS model.[8] The category variances are not free parameters, as these are estimated by the models, though somewhat differently in each case.[9]

Results

Human performance was significantly above chance for both categories: 76% of the “high” category items and 61% of the “low” category items were classified correctly. Using each participant as a distinct data point, a single sample t-test against chance performace of 50% correct produced significant results for both the high category (t_{58} = 19.6, $p < .001$) and the low one (t_{58} = 7.3, $p < .001$). The difference between the two categories was also significant: a paired samples t-test showed that people perform significantly better on the high category than the low one ($t_{58} = 8.5$, $p < .001$). This is illustrated in Figure 4, which also shows the performance of all five models. Only the RECENCY+BIAS model reproduces the human pattern. Of the other four, two of them perform below chance for the low category, and the other two are unable to correctly capture the asymmetry between the categories.

This effect is even more obvious when we plot the data from the individual subjects. Figure 5 shows the classification performance for each of the participants for both the low category (horizontal axis) and the high category (vertical axis). Most human participants fall within the triangular region, in which the high category is classified better but performance is above chance for both categories. Only the RECENCY+BIAS model falls into the same region.

In addition to looking at the data at an individual subjects level, it is useful to examine model performance at an individual stimulus level. That is, how well does each model capture the trial-by-trial changes in performance? The RECENCY+BIAS model captures 39% of the variance at the item level, compared to the 13% of the variance for the RECENCY+REGRESSION model. No other model can explain *any* of the variability in the human data at the item level.

Finally, since the quantitative evidence for the RECENCY+BIAS model is so strong, it is important to see if the the parameter estimates are sensible. The estimated

[8] Because it turns out that the model with the best performance is also the one with the most parameters, we checked that model complexity was not the source of its superior performance. This was done by calculating the Bayes factors for all models, using a uniform prior over ϕ (which varies between 0 and 1) for the three models that use exponential weighting and a uniform prior over β across the range of 0 to 10. The RECENCY+BIAS model was easily the preferred model: a numerically estimated Bayes factor produced an odds ratio of about 10^7 to 1 favoring it over the next best model.

[9] Note that including this as a free parameter would make no qualitative difference to the model predictions: all of the characteristics of the data that we focus on depend only on the classification boundary of the model, which does not depend on the variance at all.

recency parameter was $\phi = 0.25$, meaning that participants weighted each new observation x_t about one-quarter as strongly as the old prototype μ_{t-1} when updating the prototype. The bias parameter of $\beta = 4.41$ appears to be a substantial upward shift, but as Figure 4 illustrates, it is not quite large enough to overcome the "lag" induced by a reliance on old observations (as per the simpler RECENCY model). In other words, this model still lags a little behind the data.

Discussion

The experiments and computational modelling in this paper provide evidence that people adapt to changing categories in the world. This adaptation occurs not only because people weight more recent items more highly, but also because they form anticipations about the nature of the change. This work is somewhat preliminary, however, insofar as there are a number of extensions that are important to pursue. Most notable among these are the need to consider other patterns of change besides simple linear trends, and the need to extend the modelling framework to incorporate "relative judgment" approaches to categorization (Stewart, Brown, & Chater, 2002) which encode stimuli in terms of their relationship to recent items. We are currently pursuing both of these avenues.

Nevertheless, the results are interesting even without these extensions. The fact that world changes with time, and our conceptual representations must change with it, carries a number of important corollaries. In situations where the world is not constant over time, then models of category learning that assume a stationary environment are not sensible choices as rational standards for behavior. Indeed, such models do not turn out to be good models for human behavior. This does not invalidate them as good models in the contexts for which they were originally developed, but it sharply limits the generalizations one should draw from them. It is incorrect to conclude that a "rational" model that relies on an assumption of stationarity will apply to a world that changes over time and still remain a rational model in that context. Humans adapt to change, and as illustrated in the second part of the paper, if the structure of the temporal change is regular enough, we can and do learn to anticipate those changes before they occur.

That said, the best model for human performance in the task is not the ideal observer model (REGRESSION). It is instead the one that takes the simple prototype model with recency weighting (Equation 2) and makes the smallest change required to be suitable to the task. The performance of this model suggests that the adaptations that people make to anticipate systemic changes are conservative: they make smaller adjustments than the data warrant, strictly speaking (Phillips & Edwards, 1966). Conservatism itself may be a sensible adaptation (Navon, 1978), but this is not in our view the major point. The key goal is to learn what kinds of changes people are able to detect and anticipate, and how that helps us adapt to a changeable world.

Viewed from this perspective, it is notable that we were led to these results by considering the underlying inference problem that the learner has to solve: category learning in a dynamic world. In that sense, our approach offers a computational analysis, and the models that had an extrapolative component (RECENCY+BIAS, REGRESSION, and RECENCY+REGRESSION) are all inspired by that observation. However, our approach also has much in common with process-level analyses that take seriously the possibility that people must operate within the constraints imposed by memory limitations. Our work helps to bridge the two levels, bringing us closer to a united explanation of how people learn categories within a changing world.

Acknowledgements

This research was supported by ARC grant DP110104949. Salary support for DJN was provided by ARC grant DP0773794.

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.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216-233.

Brown, S. D., & Steyvers, M. (2009). Detecting and predicting changes. *Cognitive Psychology*, *58*, 49-67.

Cleveland, W. S., & Devlin, S. J. (1988). Locally-weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, *83*, 596-610.

Gökaydin, D., Ma-Wyatt, A., Navarro, D. J., & Perfors, A. (2011). Humans use different statistics for sequence analysis depending on the task. In *Proceedings of the 33rd annual conference of the cognitive science society* (p. 543-548). Austin, TX: Cognitive Science Society.

Griffiths, T. L., Sanborn, A. N., Canini, K. R., & Navarro, D. J. (2008). Categorization as nonparametric Bayesian density estimation. In M. Oaksford & N. Chater (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (p. 303-328). Oxford: Oxford University Press.

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

Luce, R. D. (1963). Detection and recognition. In *Handbook of mathematical psychology* (p. 103-189). New York: Wiley.

Navarro, D. J., & Perfors, A. (2009). Learning time-varying categories. In *Proceedings of the 31st annual conference of the cognitive science society* (p. 412-424). Austin, TX: Cognitive Science Society.

Navon, D. (1978). The importance of being conservative: Some reflections on human Bayesian behavior. *British Journal of Mathematical and Statistical Psychology*, *31*, 33-48.

Nosofsky, R. N. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 104-114.

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology*, *72*, 346-357.

Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *25*, 1161-1176.

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. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*, 1144-1167.

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 3-11.

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