Predictions with Uncertain Categorization: A Rational Model

Elizaveta Konovalova (elizaveta.konovalova@upf.edu)
Department of Economics and Business,
Universitat Pompeu Fabra, Barcelona, Spain

Gaël Le Mens (gael.le-mens@upf.edu)
Department of Economics and Business,
Universitat Pompeu Fabra, Barcelona, Spain

Abstract

A key function of categories is to help predictions about unobserved features of objects. At the same time, humans often find themselves in situations where the categories of the objects they perceive are uncertain. How do people make predictions about unobserved features in such situations? We propose a rational model that solves this problem. Our model complements existing models in that it is applicable in settings where the conditional independence assumption does not hold (features are correlated within categories) and where the features are continuous as opposed to discrete. The qualitative predictions of our model are borne out in two experiments.

Keywords: Feature inferences, Categories, Concepts, Predictions, Judgments, Rational Analysis, Bayesian Model

Introduction

According to J. Anderson, 'The basic goal of categorization is to predict the probability of various unexperienced features of objects' (Anderson, 1991). At the same time, humans often find themselves in situations where the categories of the objects they perceive are uncertain. In this article, we propose a computational model of feature prediction under uncertain categorization. We consider settings where an individual perceives some feature(s) of an object that belongs to a particular domain and makes a prediction about the value of an unobserved feature of the object. We assume that the individual has organized her knowledge of the domain into categories. We propose that the decision maker makes predictions according to a posterior distribution derived by application of Bayes' theorem. As such, our model falls into the rational analysis tradition (Anderson, 1991; Marr, 1982).

A number of prior papers have studied feature prediction under uncertain categorization (e.g., Murphy & Ross, 1994, 2010a; Griffiths, Hayes, & Newell, 2012; Papadopoulos, Hayes, & Newell, 2011). They led to interesting insights about whether and how people use categories in making predictions about unobserved features. Most of the existing studies have considered settings in which features are discrete-valued. As we explain below, a limitation of such settings is that, in this context, it is difficult to distinguish whether people do not use categories at all, or make an optimal use of the categories. By contrast, we study a setting where features are continuously-valued. In our setup, the predictions of a model that makes optimal use of the categories (our rational model) and a model that ignores categories altogether sharply differ.

Our model can be seen as an extension of the 'inference component' of Anderson's rational model of categorization (Anderson, 1991). Just as in this landmark model, the decision maker first relies on her knowledge of some feature of the object to derive the posterior probabilities that the object comes from each candidate category. Then, the decision maker uses her knowledge of the structure of each category to make predictions about the value of the unobserved feature. Anderson's model assumed that the within-category feature correlation was zero – an assumption known as *conditional independence*. Our model generalizes Anderson's model to settings where this assumption is relaxed. This extension expands the relevance of the rational model as there are many settings where it does not hold (e.g., Murphy & Ross, 2010a). In virtually all the settings where people believe that there is a causal relationship between two variables (e.g. educational achievement and income, quality and price of consumer goods), the corresponding mental representation invokes a within-category correlation (Rehder & Hastie, 2004).

Existing Paradigm

In the experimental paradigm used in the vast majority of experiments that focused on feature prediction with uncertain categorization, participants are shown a set of items of various shapes and colors divided into small number of categories, typically 4 (Murphy & Ross, 1994). Then they are told that the experimenter has a drawing of a particular shape and were asked to predict its likely color (or similar questions about the probability of an observed feature given an observed feature). An important characteristic of this paradigm is that the categories are shown graphically to the participants. The idea was to avoid complications related to memory and category learning by participants.

Suppose the two features are X and Y and there are 4 categories. Participants are asked to estimate $P(y \mid x)$, the proportion of items with $Y = y$ out of items with $X = x$. There is some evidence participants' predictions that are the same as those implied by a model that focuses on just the 'target' category, that is, the most likely category given the observed feature (Murphy & Ross, 1994). There is also some evidence that participants sometimes make predictions that are the same as those implied by a model that takes into account multiple categories (Murphy & Ross, 2010a). Still, other experiments have found evidence that participants do not pay attention to categories at all but instead are sensitive to the overall feature correlation (Hayes, Ruthven, & Newell, 2007; Papadopoulos et al., 2011; Griffiths et al., 2012).

A limitation of this paradigm pertains to the fact that the features are discrete-valued. This implies that the predictions of a model that ignores categories altogether or makes optimal use of the categories are exactly the same. This is a consequence of the law of total probability. In this case, we have

$$P(y \mid x) = \sum_{c=1}^{4} P(c \mid x) P(y \mid cx), \tag{1}$$

where $P(c \mid x)$ is the proportion of items belong to c out of all the items such that $X = x$, and $P(y \mid cx)$ is the proportion of items with $Y = y$ out of the items that both are in c and have $X = x$.

In order to estimate $P(y \mid x)$, a participant that would ignore the categories would consider all objects with $X = x$ and would respond with the proportion of objects with y among all objects with x. A participant that would consider all 4 categories would compute the proportion of items with y among the items with x *in each category* and then would compute the weighted average by multiplying each of these numbers by her estimates of $P(c \mid x)$. The responses given by the two participants would be *exactly the same*. It is therefore difficult to assess whether the participants use multiple categories (but see Murphy and Ross (2010a) for an attempt to do so using post-prediction questions). When features are continuous, however, the predictions of these two strategies differ.

A Rational Model for Predictions in Continuous Environments

By contrast to the existing paradigm, we consider a setting where the values of the two features are not discrete, but real-valued random variables X (first feature) and Y (second feature). We assume the individual has organized her knowledge of the domain of objects in a set of categories $\mathcal{C}$. Following recent work, we model mental categories using probability distribution functions (*pdfs*) on the feature space (Ashby & Alfonso-Reese, 1995; Sanborn, Griffiths, & Shiffrin, 2010). Let $c \in \mathcal{C}$ be a category. We denote by $f_c(x,y)$ the value of the associated *pdf* at position (x,y) in the feature space, where x denotes the value of the first feature and y denotes the value of the second feature. This *pdf* denotes the prior belief of the individual over positions given that she knows that an object is from category c.

Now suppose that the individual observes that the first feature has value x and predicts the value of the second, unobserved feature. We assume her predictions are driven by her posterior on the value of the second feature given her observation of the first value of the first feature: $f(y \mid x)$. How does the individual compute this quantity, assuming that she does not have a pre-existing mental representation of the (probabilistic) relation between x and y?

We propose that the individual relies on her mental representation of the categories to make the prediction. That is, she will make use of the category *pdfs* she has in memory. More precisely, we propose that the individual's posterior belief on the value of the second feature is a weighted sum of the posteriors obtained for each possible category:

$$f(y \mid x) = \sum_{c \in \mathcal{C}} p(c \mid x) f_c(y \mid x), \tag{2}$$

where $p(c \mid x)$ is the subjective probability that the object comes from category c given the observed feature value x on the first dimension and $f_c(y \mid x)$ is the marginal distribution of value of the second feature, conditional on the fact that the object is in category c and that its first feature has value x.

This model is realistic to the extent the agent can compute the components of the RHS on the basis of her mental representations. Here, we assume she does so by applying the rules of probability calculus. First, consider the posterior distribution of Y given x and c. We have:

$$f_c(y \mid x) = \frac{f_c(x,y)}{\int_v f_c(x,v)dv}, \tag{3}$$

Second, we assume the agent also computes the probabilities that the item comes from each candidate category in a way that is consistent with Bayes' theorem. We have

$$p(c \mid x) = \frac{P(c) f_c(x)}{f(x)} = \frac{P(c) \int_v f_c(x,v)dv}{\sum_{c \in \mathcal{C}} P(c) \int_v f_c(x,v)dv}, \tag{4}$$

where $P(c)$ is the prior on the category. This is the probability that an object about which the individual has no information comes from category c. In the category learning literature, this term is frequently called the 'category bias'.

The predictions of our model follow the rules of probability calculus. Thus, our model makes rational predictions (given the constraints imposed by the mental representation of the categories). Next, we illustrate how the model works by analyzing what happens when the category *pdfs* are bi-variate normal distributions.

Suppose we have two categories ($\mathcal{C} = \{1,2\}$) and that categories can be represented by bi-variate normal distributions as follows:

$$\begin{pmatrix} X_c \\ Y_c \end{pmatrix} \sim N\Big(\begin{pmatrix} \mu_{xc} \\ \mu_{yc} \end{pmatrix}; \begin{pmatrix} \sigma_{xc}^2 & \rho_c \sigma_{xc} \sigma_{yc} \\ \rho_c \sigma_{xc} \sigma_{yc} & \sigma_{yc}^2 \end{pmatrix} \Big), \tag{5}$$

where μ_{xc} and μ_{yc} are the category means on the two features, σ_{xc} and σ_{yc} are the standard deviations on the two features and ρ_c is the within-category correlation for category c. We first assume that there is no within-category correlation, consistent with the conditional independence assumption (Anderson, 1991). Then we consider the general case.

Model Predictions - with Conditional Independence

Assuming conditional independence amounts to assuming $\rho_{c1} = \rho_{c2} = 0$. Some algebra leads to

$$f(y \mid x) = p(c_1 \mid x) f_{\mu_{y1}, \sigma_{y1}}(y) + p(c_2 \mid x) f_{\mu_{y2}, \sigma_{y2}}(y), \tag{6}$$

where f_{μ_y, σ_y} denotes the density of a normal distribution with mean μ_y and standard deviation σ_y, $p(c_2 \mid x) = 1 - p(c_1 \mid x)$, and

$$p(c_1 \mid x) = \frac{1}{1 + e^{ax^2 - bx + c}}, \tag{7}$$

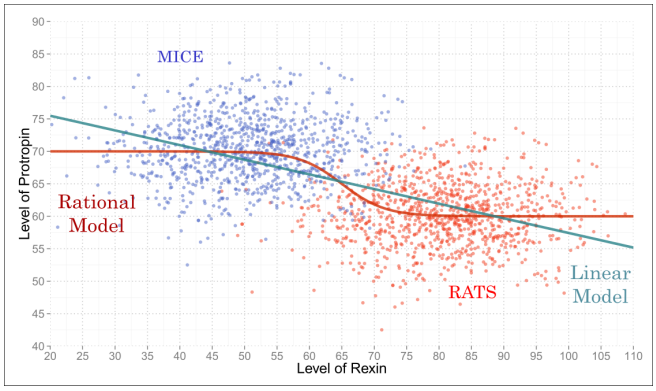

Figure 1: Categories used in Experiment 1. The solid lines represent the mean of the posterior implied by our model and the linear model. Participants were shown the level of 'Rexin' (x-axis) and were asked to predict the level of 'Protropin' (y-axis). See column 'True' in Table 1 for parameter values.

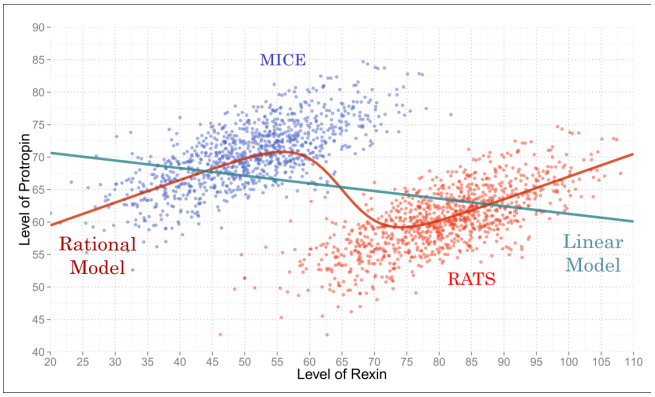

Figure 2: Categories used in Experiment 2. The solid lines represent the mean of the posterior implied by our model and the linear model. The parameters are the same as for Figure 1, except for the within-category correlations: $\rho_R = \rho_M = 0.7$.

with

$$a = \frac{\sigma_{x2}^2 - \sigma_{x1}^2}{2\sigma_{x2}^2\sigma_{x1}^2}, b = \frac{\sigma_{x2}^2\mu_{x1} - \sigma_{x1}^2\mu_{x2}}{\sigma_{x2}^2\sigma_{x1}^2}, c = \frac{\sigma_{x2}^2\mu_{x1}^2 - \sigma_{x1}^2\mu_{x2}^2}{2\sigma_{x2}^2\sigma_{x1}^2} + log\frac{\sigma_{x2}}{\sigma_{x1}}.$$

See Figure 1 for an illustration. The predictions made by the model are sensitive to the relative positions of the categories and make a 'smooth' transition from one category to the other. Due to the fact that it is essentially a version of Anderson's rational model ('AM'), we will refer to this model as 'Anderson's model' in subsequent discussions. Its predictions are different from the predictions of other existing models.

Ignoring the Categories: Linear Model (LM) A simple model that would ignore categories altogether would make predictions according to a regression line with negative slope (see Figure 1 for illustration).

Single Category - Independent Features (SCI) We refer to the most likely category given the observed feature (x) as the 'target' category (this is category 1 if $p(c_1 \mid x) > .5$, as per eq. 7). This is the same as the rational model, but with all the weight on the target category (c^*). In this case, $f(y \mid x) = f_{c^*}(y \mid x)$, where $f_{c^*} = f_{\mu_{y1},\sigma_{y1}}$ if the target category is category 1, and $f_{c^*} = f_{\mu_{y2},\sigma_{y2}}$ otherwise. The mean of the posterior implied by this model follows a 'step function' where the two steps are at $y = \mu_{y1}$ and $y = \mu_{y2}$ and the jump is situated where x is such that $p(c_1 \mid x) = .5$ (with the experimental parameters, this is obtained for $x = 65$).

Model Predictions - General Case

When the conditional independence assumption does not hold, the posterior is given by

$$f(y \mid x) = p(c_1 \mid x) f_{\mu_{yc_1} + \frac{\sigma_{yc_1}}{\sigma_{xc_1}}\rho_{c_1}(x - \mu_{xc_1}), \sigma_{y1}}(y) + p(c_2 \mid x) f_{\mu_{yc_2} + \frac{\sigma_{yc_2}}{\sigma_{xc_2}}\rho_{c_2}(x - \mu_{xc_2}), \sigma_{y2}}(y), \quad (8)$$

where $p(c_1 \mid x)$ is given by the same equation as before (equation 7). See Figure 2 for an illustration. In this case, the prediction of the second feature is influenced by the positions of the categories as well as by the internal structure of the categories (the within-category correlation between X and Y). We will refer to this model as the rational model with possible correlation (RMC).

In the setting of the Figure 2, the mean of the posterior implied by the linear model would be a downward slopping line. The mean of the posterior implied by the Single Category - Independent Features model would be a step function, just as before. Another relevant comparison model is a model that uses just the target category but is sensitive to within-category feature correlations.

Single Category - Correlated Features (SCC) Let c^* be the most likely category given the observation of the first feature. We have $f(y \mid x) = f_{c^*}(y \mid x)$, where $f_{c^*} = f_{\mu_{yc_1} + \frac{\sigma_{yc_1}}{\sigma_{xc_1}}\rho_{c_1}(x - \mu_{xc_1}), \sigma_{y1}}$ if the target category is category 1, and $f_{c^*} = f_{\mu_{yc_2} + \frac{\sigma_{yc_2}}{\sigma_{xc_2}}\rho_{c_2}(x - \mu_{xc_2}), \sigma_{y2}}$ otherwise.

Experiment 1: With Conditional Independence

Participants faced a feature prediction task that closely matches the setting of the previous section. They learnt two categories in a two dimensional feature space and then made a series of predictions about the value of the second feature on the basis of the value of the first feature of an item.

Design

To avoid the influence of unobserved prior knowledge, our experiment used artificial categories. Participants had to assume they were biochemists who studied the levels of two hormones in blood samples coming from two categories of animals. (see Kemp, Shafto, and Tenenbaum (2012) for a similar setup). The hormones were called 'Rexin' and 'Pro-

tropin' and the two categories of animals were 'Mice' and 'Rats.' Similarly to prior literature on feature prediction with uncertain categorization and in order to avoid issues related to memory, we provided the participants with visual representations of the categories in the form of scatter plots of the exemplars of the two categories. The data was generated on the basis of eq. 5 and the parameters in the 'True' column of Table 1 (see Figure 1).

The flow of the experiment was as follows: After reading general instructions, participants were told their lab had a collection of Rat blood samples. They were shown a scatter plot of the levels of Rexin and Protropin in these blood samples and given as much time as they wanted to study it. Then they were asked to make a series of 13 successive predictions of the likely level of Protropin given the level of Rexin found in a blood sample (the scatter plot was visible on the computer screen while participants made the predictions). We asked participants to make these predictions so that they would become familiar with the relation between the levels of Rexin and Protropin (a positive within-category correlation). No feedback was provided about the predictions. Then participants went through a similar procedure for the Mouse blood samples. To conclude this learning stage, participants were shown a graph with the Rat sample data and Mouse sample data (the scatter plots of Figure 1, without the prediction of the rational model).

In the next stage, participants were told that a batch of new blood sample had just arrived at their lab and that these blood samples had already been tested for Rexin. They were also told that the 'label on the blood sample has been erased and thus you do not know if it belongs to a rat or a mouse.' Participants were asked to predict the likely level of Protropin for 48 blood samples. The question was 'What is the likely level of Protropin in this blood sample?'. Participants answered using a slider scale with minimal value 40, maximal value 90, and increments of 1 unit.

Participants

29 participants were recruited via Amazon Mechanical Turk. 1 participant was eliminated due to a technical error that occurred during the experiment.

Results

Models were estimated on a participant-by-participant basis and evaluated in terms of the BIC criterion (minus the log-likelihood minus a penalty increasing in the number of free parameters). Table 1 reports the mean estimated parameter values (across participants). They are close to the true parameter value, which suggests that collectively, participants understood the task and behaved in a way generally consistent with the predictions of the rational model.

We proceed to two model comparisons (see Table 2). In the first comparison, we compare Anderson's rational model (AM), the single category independent feature model (SCI) and the linear model (LM). Anderson's model provides the best fit for 64% of the participants, whereas the two other models provide the best fit for just 18% of the participants. This suggests that most participants generally took into account the two categories when predicting the value of the second feature. Also, they displayed the "smooth" transition between the categories predicted by Anderson's model.

Table 1: Estimated model parameters. Parameters were estimated separately for each participant. The values are the mean estimated parameters across participants. AM: Anderson's rational model, SCI: the single category independent feature model; RMC: rational model with possible within-category feature correlation; SCC: single category model with possible within-category feature correlation; LM: linear model.

Experiment 1						
Param.	**True**	**AM**	**SCI**	**RMC**	**SCC**	**LM**
$\mu_{x,R}$	80	78.5	77.5	81.5	82.1	
$\mu_{y,R}$	60	60.4	60.4	60.9	61.0	
$\mu_{x,M}$	50	48.2	47.5	49.2	52.0	
$\mu_{y,R}$	70	68.8	69.2	68.6	68.1	
σ_x	10	7.1	7.4	7.5	12.5	
σ_y	5	0.2	1.7	2.5	3.3	
$\rho_R = \rho_M$	0	NA	NA	-0.1	-0.2	
α	78					74.7
β	-.2					-0.2
σ^2	5.7					4.1
BIC	NA	233.3	257.7	231.5	260.3	273.6
Experiment 2						
$\mu_{x,R}$	80	79.3	79.9	79.6	80.1	
$\mu_{y,R}$	60	60.8	60.9	61.2	60.8	
$\mu_{x,M}$	50	47.4	49.9	47.1	50.4	
$\mu_{y,R}$	70	69.3	69.1	68.7	68.7	
σ_x	10	3.3	9.9	5.8	10.1	
σ_y	5	5.8	5.0	3.0	4.9	
$\rho_R = \rho_M$	0.7	NA	NA	0.5	0.5	
α	73					66.8
β	-.12					-0.03
σ^2	6.7					6.7
BIC	NA	316.1	317.1	246.2	285.9	326.6

In the second comparison we included two additional models: a version of the rational model with possible within-category feature correlation (RMC) and a version of the single category model with possible within-category feature correlation (SCC). If people behave rationally, according to the task environment, these models should perform more poorly than their equivalents with 0 within-category feature correlation. This is because the true correlation is 0, and these models have the correlation coefficient as one more free parameter. They should suffer some penalty in term of the BIC. We find that Anderson's model provides the best fit for 36% of the participants, about half as many as in the previous comparison. The rational model with within-category correlation (RMC) provides the best fit for 46% of the participants. The other models are the best fitting model for very few participants. These numbers suggest that a number of participants behaved as if there were some within-category feature corre-

lation despite the fact there was none. This could be because it is hard for people to grasp the concept of randomness or the absence of a pattern (Nickerson, 2002). Moreover, Grice's maxim of quantity suggests that participants might not expect the experimenter to show a graph that communicates an absence of relation (Grice, 1975).

Despite this pattern of behavior, this analysis suggests that most participants considered the two categories, because the two models with smooth transitions between categories (AM and RMC) provide a much better fit than the models that focus on a single category of the linear model (LM). Next, we adapt this design to a setting with positive within-category correlation. In this case, our rational model (RMC) and Anderson's model (AM) make distinct predictions.

Experiment 2: With Positive Within-Category Correlation

In addition to testing our model in a setting without conditional independence, we wanted to see whether we could manipulate the propensity of participants to rely on just the target category or the two candidate categories. Many of the studies that had found that participants tend to rely just on the target category included a question that asked participants about the most likely category of the stimulus before making their predictions. There is evidence that the wording of this question affects the propensity to rely on one or multiple categories when making predictions (Murphy & Ross, 2010b; Murphy, Chen, & Ross, 2012; Hayes & Newell, 2009). We included a similar manipulation in our study.

Design

The design of the experiment was the same as in Experiment 1, but with a within-category feature correlation of .7 (see Figure 2). There were 3 conditions. In the control condition, participants were just asked to predict the second feature value upon seeing the value fo the first feature, as in Experiment 1. In the 'MC condition', participants were asked about the most likely category before predicting the value of the second feature. This was multiple choice question: 'From which animal did the blood sample come from?'. The choices were 'Mouse' and 'Rat'. In the 'SL condition', participants were asked the same question, but answered using a continuous slider that went from 'Definitely a Mouse' (left) to 'Possibility a rat or a mouse' (middle) to 'Definitely a Rat' (right). We predicted that the SL condition would make people more aware of the uncertainty about the category of the item and thus increase their propensity to rely on two categories, at least as compared to what happens in the MC condition.

Table 2: Percentage of participants whose feature predictions were best fit by each of the candidate models.

	Experiment 1 Comparison		**Experiment 2** Condition		
Model	(1)	(2)	MC	SL	Control
AM: Anderson	64%	36%	0%	0%	4%
SCI: Single Cat. Indep. Features	18%	4%	0%	0%	0%
SCC: Single Cat. Corr. Features		0%	17%	24%	25%
RMC: Rat. Mod. Corr. Features		46%	72%	69%	63%
LM: Linear	18%	14%	10%	7%	8%
Nb part.	28	28	29	29	24

Participants

102 participants were recruited via Amazon Turk. 20 participants were eliminated from the analysis because their responses seemed very inconsistent with the stimuli.[1]

Results

We fitted the 5 candidate models on a participant-by-participant basis and compared them in terms of the BIC criterion (see Tables 1 & 2). The rational model (RMC) provided the best fit to the data in all 3 conditions (it is the best fitting model for 60 to 70% of the participants). For about 20% of the participants, the best fit is a model that focuses on the most likely category (SCC or SCI).

Comparisons of the percentage of participants for whom the best fit is the rational model or a single category model do not show meaningful differences across categories. In order to uncover differences, we estimated a quasi-rational model with a free parameter that characterizes the propensity to rely on multiple categories. This model assumes that the posterior is given by

$$f(y \mid x) = p(c_1 \mid x)^{\alpha} f_{\mu_{yc_1}+\frac{\sigma_{yc_1}}{\sigma_{xc_1}}\rho_{c_1}(x-\mu_{xc_1}),\sigma_{y1}}(y) + p(c_2 \mid x)^{\alpha} f_{\mu_{yc_2}+\frac{\sigma_{yc_2}}{\sigma_{xc_2}}\rho_{c_2}(x-\mu_{xc_2}),\sigma_{y2}}(y), \quad (9)$$

where $\alpha > 0$. When $\alpha = 1$, the model reduces to the rational model (eq. 8). When α is high, the model becomes close to the single-category correlated feature model (SCC), and when α is close to 0, the model gives about equal weight to both categories, irrespective of the value of the observed feature.

Maximum Likelihood Estimations on a participant-by-participant basis give the following proportions of participants with α higher than 1: MC condition: 83%, SL condition: SL: 62% and control condition: 71%. These proportions are all significantly higher than 50% but not significantly different from each others (one-sided binomial tests with level of .05). Although the differences are not large, the ranking of the three proportions is consistent with our expectations. Maybe more significantly, the fact that α is higher than 1 for

[1] For each participant, we regressed the Y values (the predictions) on X (the values of the first feature, shown to them). Those participants for whom the regression coefficient was significantly positive were dropped from the analysis. The reason is that the rational model can fit such pattern of predictions well with $\mu_{y,R} > \mu_{y,M}$. But such pattern can hardly be considered rational given given the true values of $\mu_{y,R}$ and $\mu_{y,M}$ ($\mu_{y,R} < \mu_{y,M}$ - see Figure 2). Ancillary analyses with all the participants lead to similar results.

most participants in all conditions suggests that most participants gave more weight to the target category than what is prescribed by the rational model.

Discussion

Model comparisons suggest that the rational model provides an appropriate characterization of the behavior of a large proportion of the participants, when compared to other models. This implies that most participants consider the two candidate categories when making predictions about the unobserved feature. This might seem surprising in light of the existing evidence gathered by Murphy, Ross and colleagues that the majority of participants tend to rely on just the target category (in their experiments about 25% of the participants rely on multiple categories). But our analyses with the quasi-rational model suggest that most participants in fact give too much weight to the target category, at least compared to the prescription of the rational model. Seen with this lens, our results are not inconsistent with the prior findings, but rather refine them and extend these to a different experimental paradigm.

Discussion & Conclusion

Our rational model implies that when the category of the item is uncertain, participants should give some weight to the predictions implied by membership in the two candidate categories. This should be the case both under conditional independence or when there is within-category feature correlation. Our empirical results suggest that a majority of participants behaved according to this qualitative prediction. At the same time, most but not all participants gave too much weight to the most likely category. This is broadly in line with prior empirical findings in the literature on category-based feature prediction.

Our model is a computational model and, as such, it does not specify how people might perform the computations that lead to these predictions (Marr, 1982). Nosofsky (2015) recently proposed an algorithmic model that achieves such predictions when the features are discrete-valued. Adapting this exemplar model to the case where feature values are continuous is an interesting avenue for further research.

Finally, a potential limitation of our experiments is that people were watching the data (the scatter plots) when making the predictions. A possible interpretation of our findings is thus that people engaged in some elaborate form of curve fitting on the basis of what they were looking at. A natural next step is to run similar experiments where participants first learn the categories and then make feature predictions on the basis of memorized categories.

Acknowledgments

We appreciate the discussion with participants in the Behavioral and Management Breakfast at UPF, the ConCats seminar at NYU and comments by Mike Hannan and Robin Hogarth. This research was funded by a GSE Seed Grant, MINECO grants #PSI2013-41909-P and #RYC-2014-15035 to Gaël Le Mens.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*(2), 216–233.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics speech arts* (pp. 41–58). New York.

Griffiths, O., Hayes, B. K., & Newell, B. R. (2012). Feature-based versus category-based induction with uncertain categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 576–595.

Hayes, B. K., & Newell, B. R. (2009). Induction with uncertain categories: When do people consider the category alternatives? *Memory & Cognition*, *37*(6), 730–743.

Hayes, B. K., Ruthven, C., & Newell, B. R. (2007). Inferring properties when categorization is uncertain: A feature-conjunction account. In *Proceedings of the 29th annual conference of the cognitive science society* (pp. 209–214).

Kemp, C., Shafto, P., & Tenenbaum, J. B. (2012). An integrated account of generalization across objects and features. *Cognitive Psychology*, *64*(1), 35–73.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. NewYork: Freeman.

Murphy, G. L., Chen, S. Y., & Ross, B. H. (2012). Reasoning with uncertain categories. *Thinking & Reasoning*, *18*(1), 81–117.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive psychology*, *27*(2), 148–193.

Murphy, G. L., & Ross, B. H. (2010a). Category vs. object knowledge in category-based induction. *Journal of Memory and Language*, *63*(1), 1–17.

Murphy, G. L., & Ross, B. H. (2010b). Uncertainty in category-based induction: When do people integrate across categories? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 263–276.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological review*, *109*(2), 330–357.

Nosofsky, R. M. (2015). An exemplar-model account of feature inference from uncertain categorizations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1929–1941.

Papadopoulos, C., Hayes, B. K., & Newell, B. R. (2011). Noncategorical approaches to feature prediction with uncertain categories. *Memory & cognition*, *39*(2), 304–318.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition*, *91*(2), 113–153.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010, March). Uncovering mental representations with markov chain monte carlo. *Cognitive Psychology*, *60*(2), 63–106.