Further attempts to clarify the importance of category variability for categorisation

Amotz Perlman [a] , Ulrike Hahn [b] , Darren J. Edwards [c] & Emmanuel M. Pothos [c]

[a] Ben-Gurion University of the Negev, Beer Sheva, Israel

[b] Cardiff University, Cardiff, UK

[c] Swansea University, Swansea, UK

Version of record first published: 07 Nov 2011.

PLEASE SCROLL DOWN FOR ARTICLE

Further attempts to clarify the importance of category variability for categorisation

Amotz Perlman[1], Ulrike Hahn[2], Darren J. Edwards[3], and Emmanuel M. Pothos[3]

[1]Ben-Gurion University of the Negev, Beer Sheva, Israel
[2]Cardiff University, Cardiff, UK
[3]Swansea University, Swansea, UK

The issue of how category variability affects classification of novel instances is an important one for assessing theories of categorisation, yet previous research cannot provide a compelling conclusion. In five experiments we reexamine some of the factors thought to affect participant performance. In Experiments 1 and 2, participants almost always classified the test item as belonging to the high variability category. By contrast, in Experiment 3 we employed an alternative experimental paradigm, where the difference in variability of the two categories was less salient. In that case, participants tended to classify a test item as belonging to the low variability category. Two additional experiments (4 and 5) explored in detail the differences between Experiments 1, 2 on the one hand, and 3 on the other. Some insight into the underlying psychological processes can be provided by computational models of categorisation, and we focus on the continuous version of Anderson's (1991) Rational Model, which has not been explored before in this context. The model predicts that test instances exactly halfway between the prototypes of two categories should be classified into the more variable category, consistent with the bulk of empirical findings. We also provided a comparison with a slightly reduced version of the Generalised Context Model (GCM) to show that its predictions are consistent with those from the Rational Model, for our stimulus sets.

Interitem variability affects behaviour in a number of related areas. In categorisation, the problem is how category variability influences item classification. This is an important issue both because previous research has led to somewhat conflicting findings and because computational models of categorisation make strong predictions. We discuss some of the general literature on variability and proceed to present the relevant data from categorisation, which is the focus of the present study, and corresponding computational analyses.

Variability plays a central role in studies of human inductive inference, where it has been argued that, all other things being equal, more variable, diverse evidence should give rise to stronger inductive arguments. This diversity principle has been highlighted in the philosophy of science (see for example, Bacon, 1898/1620; Carnap, 1950; Horwich, 1982; Howson & Urbach, 1993; Nagel, 1939) and there has been considerable experimental work examining the extent to which it is adhered to in our everyday judgements

by both adults (Lopez, 1995; Lopez, Atran, Coley, Medin, & Schaffer, 1997; Osherson, Smith, Wilkie, Lopez, & Shafir, 1990) and children (Carey, 1985; Gutheil & Gelman, 1997; Heit & Hahn, 2001; Lopez, Gelman, Gutheil, & Smith, 1992).

The influence of diversity on inductive reasoning is likely to be related to the influence of diversity or variability on categorisation, although this link is seldom made. There is evidence that infants are sensitive to variability of categories (e.g., Mareschal, Quinn, & French, 2002; Quinn, Eimas, & Rosenkrantz, 1993; Younger, 1985). With adults, it has been repeatedly reported that more variable observations promote broader or stronger generalisations (e.g., Cohen, Nosofsky, & Zaki, 2001; Fried & Holyoak, 1984; Homa & Vosburgh, 1976; Posner & Keele, 1968; Rips, 1989; Stewart & Chater, 2002)—we call this effect the category variability effect. At the same time, there is some evidence that more variable categories are harder to acquire (e.g., Fried & Holyoak, 1984; Peterson, Meagher, Chait, & Gillie, 1973; Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968; Pothos, Edwards, & Perlman, 2011). Finally, there has been some investigation of the effects of interitem variability on memory (e.g., Homa & Vosburgh, 1976). Moreover, these various effects can be linked: Hahn, Bailey, and Elvin (2005) observed effects of category variability on learning, category generalisation, and old/new recognition that could be systematically related through the exemplar-based framework of the Generalised Context Model (GCM; e.g., Nosofsky, 1986, 1988). Detailed modelling revealed that variability systematically affected sensitivity to distance in psychological space, and hence, similarity across these tasks.

Our focus is the category variability effect with adults. Closer scrutiny of the literature reveals that, although there is plenty of empirical evidence regarding the impact of category variability on classification, results are not clear cut and classification performance seems heavily influenced by minor, and seemingly superficial, methodological variations. Consequently, the main objective of the present paper is to identify a number of pertinent methodological manipulations and explore them further. A secondary objective is to examine whether an influential categorisation model, Anderson's (1991) Rational Model, can provide any insight into the underlying computational process. This model has not been explored before with respect to this issue, despite its clear relevance.

A well-known experiment by Rips (1989) provides a graphic illustration. Rips gave participants a sparse description of an object, such as "a circular object with a 3-in. diameter", and asked them to visualise it. He next asked one group of participants if it was more *similar* to a pizza or a particular coin (a US currency quarter). Most of the participants said that it was more similar to the quarter; presumably, the reason was that its diameter was closer to the diameter of the average quarter than to that of the average pizza. He then asked a second group of participants if it was more *likely* to be a pizza or a quarter. Most of the participants decided pizza; seemingly because pizzas are found in a range of diameters (i.e., they are the more variable category); hence, it is more likely that they could include an exemplar with a 3-inch diameter as well. By contrast, because the range of diameters in the coin category is so restricted, it seems unlikely that there would be exemplars of the category that deviate from the typically observed extremes. An earlier report of a similar effect, but with meaningless stimuli, was made by Fried and Holyoak (1984), who showed that participants classified checkerboard examples that were closer to the prototype (mean) of a lower variability category, as belonging to a high variability category, which was further away.

There have been a number of extensions to Rips's (1989) experiment, which illustrate some subtle methodological issues. Cohen et al. (2001) replicated Rips's finding in two ways. First, they employed one-dimensional stimuli (vertical lines), organised into a low and a high variability category, with a test instance halfway between the nearest neighbours of the two categories (in the following, test instances halfway between the nearest neighbours of two categories will be referred to as NN-halfway instances, to distinguish them from test instances located halfway between the prototypes of the two trained categories which will be referred to as P-halfway). Again, it was found that increasing the variability of the high variability category favoured classification into the high variability category (note though that the classification rate for the high variability category was between 29.5% and 47.1%). Second, Cohen et al. also examined stimuli that were colour patches varying along two dimensions, brightness and hue (Exp. 2). In four conditions the NN-halfway test instance was always numerically more likely to be classified into the high variability category (though these trends were not always significant).

Cohen et al. (2001) provided some interesting manipulations regarding the importance of category variability in classification. Two methodological issues, in particular, are worth highlighting. First, in their Experiment 2, Cohen et al. defined the low and the high variability category in such a way that, if participants were to classify the test instance in the low variability category, then there would not always be a linear category boundary separating the exemplars of the two categories. By contrast, if participants classified the test instance in the high variability category (as most of them did), then there was always a simple linear boundary separating the exemplars of the two categories (this point is immediately obvious by inspecting Figure 4 in Cohen et al.'s paper). There is evidence that whether a category boundary is linearly separable or not affects performance with the corresponding classification (e.g., Blair & Homa, 2001; Ruts, Storms, & Hampton, 2004). Second, in Experiment 1, the presentation frequency of the exemplars favoured the nearest neighbours of the two categories. Given the simple, schematic nature of the stimuli, it is possible that this may have influenced the kind of categories participants extracted and perhaps distorted the representation of category variability. Cohen et al. had good reasons for adopting this design, but we wanted to study the issue of category variability on classification without these methodological complications.

A related study is that of Sakamoto, Jones, and Love (2008), where participants saw test items that were NN-halfway; again, these test items were more likely to be classified into the more variable category. In Sakamoto et al.'s study the stimuli were lines, in a way analogous to that of Cohen et al. (2001), and the training procedure was that of a standard supervised categorisation experiment (participants went through several presentations of the training stimuli, each time trying to guess their category membership and receiving corrective feedback). Sakamoto et al.'s results are consistent with those of Cohen et al., but are somewhat at odds with earlier results from Stewart and Chater's (2002) Experiment 1. Stewart and Chater (in their Exp. 1) also considered NN-halfway test instances, but manipulated format of presentation and the presence of a hint regarding category variability. When Stewart and Chater employed a sequential mode of stimulus presentation (in their Exp. 1), the probability of classification into the high variability category (without a corresponding hint) was only .25 (it was .37 with the hint). However, with simultaneous presentation of all the stimuli, this probability rose to .51 without a hint to participants that one category was more variable and .74 with such a hint. Thus, when it comes to sequential presentation of the stimuli, and NN-halfway test instances, there is some conflict in the literature.

Overall, more recent research has somewhat emphasised NN-halfway test instances. Although there are important theoretical reasons for studying NN-halfway test instances, it is worth noting that an NN-halfway item will actually be closer to the prototype of the less variable category, compared to an item that is halfway between the prototypes of two categories (a P-halfway item), as illustrated in Figure 1. Indeed, Cohen et al. (2001) found that a slightly reduced version of the GCM could not account for the bias to classify NN-halfway instances into the more variable category. But, it is also theoretically pertinent to study P-halfway instances (cf. Hahn et al., 2005). One reason is that the nearest neighbours between two categories may not be as important an aspect of category representation as, by contrast, the prototypes (note that in the Cohen et al. study the salience of the nearest neighbours was artificially enhanced in some conditions, by increasing the frequency of these items). Thus, we think that the literature on category variability could be usefully extended by a focused study involving P-halfway instances and this is what we aim to achieve with the present research (taking into account a range of relevant methodological manipulations). Notably, some work indicates that the default classification bias is to place an NN-halfway exemplar into the high variability category (Cohen et al., 2001; Sakamoto et al., 2008), others that if category variability is not made salient it will be classified in the low variability category (Stewart & Chater, 2002). There are different ways in which variability can be made salient and this issue will be considered in our experiments.

Our understanding of empirical results can be informed by computational models of categorisation. Indeed, many computational models of categorisation are sensitive to the underlying distributional characteristics of the categories that are learned. So, what do they predict regarding P-halfway or NN-halfway instances? Note first that, as baseline prototype models determine the classification of novel instances only on the basis of category prototypes, they are blind to the effects
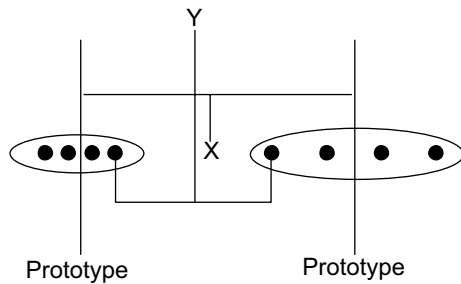
Figure 1.   Point X is halfway between the prototypes of the two categories, and Point Y indicates the point halfway between the nearest neighbours of the two categories. It can be seen that Point Y is closer to the prototype of the less variable category.

of category variability (but see Vanpaemel & Storms, 2008, for extensions of prototype models). Exemplar models are, in principle, sensitive to category variability, since category representations typically involve information about all available exemplars. Cohen et al. (2001); see also Stewart & Chater, 2002) examined a version of the GCM with only one free parameter (the sensitivity parameter) and showed that the model was more consistent with classification of NN-halfway test instances into the *low* variability categories. But, as noted, the conclusions from this examination probably also depend on differences in relative frequency between category exemplars and the complexity of the category boundaries.

Stewart and Chater (2002) examined a "distributional" approach (from Ashby & Townsend, 1986), according to which probability distributions are used to represent categories and these distributions are fitted using the observed stimuli (see also Fried & Holyoak, 1984). On this account classification of a new exemplar is based on the relative likelihood of belonging to the distribution of each of the possible categories. An NN- halfway test item should typically be classified into the high variability category, because the tight bunching of the exemplars in the low variability category means that the test item is more standard deviations away from the mean of that category. A similar analysis has recently been reported by Hsu and Griffiths (2010), who examined the predictions of a baseline Bayesian model of classification, based on an assumption of normally distributed properties. They also showed that their model favours classification into the more variable category.

Stewart and Chater's (2002) analysis of the distributional approach and Hsu and Griffiths' (2010) examination of the baseline Bayesian model suggest that any categorisation model that represents categories in terms of the probability distribution of their members ought to predict that a halfway instance would be classified in the more variable category. An important such model is Anderson's (1991); see also Sanborn, Griffiths, & Navarro, 2010) Rational Model, which has come to be the most widely considered Bayesian model of categorisation.

Although one would expect an application of the Rational Model to the problem of category variability to lead to conclusions consistent with those of Stewart and Chater (2002), it is an important exercise to verify that this is indeed the case. Therefore, a secondary purpose of this work was to apply the continuous version of Anderson's (1991); Anderson & Fincham, 1996) Rational Model to the problem of how category variability affects the classification of novel stimuli. Note that the continuous version of the Rational Model (which assumes items are represented with continuous dimensions) has been examined considerably less than the better known discrete version (which assumes that items are represented with discrete features).

The Rational Model assumes that the process of categorisation is one of Bayesian inference for novel stimuli. The key explanatory components of the Rational Model are the assumptions it makes regarding how experience is represented and employed when categorising novel instances. Specifically, it assumes that naïve observers base their category extensions on knowledge of the underlying category distributions (i.e., distributions of category exemplars). Thus, when a test instance is consistent with a category distribution, it is more likely to be classified into the category. The Rational Model clearly predicts that classification would favour the more variable category. Thus, the issue of whether participants are sensitive to category variability or not bears on whether participants can be assumed to encode distributional information from the observed instances and employ it for classification in the way prescribed by the Rational Model.

More generally, examination of the Rational Model adds to the effort of understanding cognitive process within formal probabilistic frameworks (e.g., Busemeyer, Pothos, Franco, & Trueblood, 2011; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Oaksford & Chater, 2007;

Tenenbaum, Griffiths, & Kemp, 2006; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). Finally, it is worth noting that there are other categorisation models which could be applied to the category variability issue, such as the simplicity model (Pothos & Chater, 2002) or SUSTAIN (Love, Medin, & Gureckis, 2004). Our choice to study the Rational Model relates primarily to the fact that, as we shall see, this model provides a particularly clear intuition as to why category variability should matter.

THE CONTINUOUS VERSION OF THE RATIONAL MODEL

Anderson's (1991) Rational Model is an incremental, Bayesian model of unsupervised categorisation. It assigns a new stimulus with feature structure F to whichever category k makes F most probable. For example, a new object that looks like a "cat" would be assigned to the category of cats, since the feature structure of the object is most probable given this category membership. Specifically, the probability of classification of a novel instance into category k is given by the product $P(k) P(F|k)$. $P(k) = \frac{cn_k}{(1-c)+cn}$, whereby n_k is the number of stimuli assigned to category k so far, n is the total number of classified stimuli, and c is the coupling parameter. The coupling parameter determines how likely it is that a new instance will be assigned to a new category. Thus, c indirectly determines the number of categories that the Rational Model will produce for a stimulus set.

The probability that the new object comes from a new category is given by $P(0) = \frac{1-c}{(1-c)+cn}$. $P(F|k)$ is computed as $\prod_i f_i(x|k)$, where i indexes the different dimensions of variation of the stimuli and x indicates the different values dimension i can take. That is, $f_i(x|k)$ is the probability of displaying value x on dimension i in category k, and is approximated by $t_{a_i}(\mu_i, \sigma_i\sqrt{1+1/\lambda_i})$, which is the t distribution with a_i degrees of freedom. μ_i and σ_i^2 are given by $\frac{\lambda_0\mu_0+n\bar{y}}{\lambda_0+n}$ and $\frac{\alpha_0\sigma_0^2+(n-1)s^2+\frac{\lambda_0 n}{\lambda_0+n}(\mu_0-\bar{y})^2}{\alpha_0+n}$, respectively, whereby $\lambda_i = \lambda_0 + n$, $\alpha_i = \alpha_0 + n$, n is the number of observations in category k, y is their mean, and s^2 is their variance. Finally, $\alpha_0 = 1 = \lambda_0$, μ_0, is the halfway point of the range of all instances and is the square of a quarter of the range. This

particular form for the Rational Model deviates very slightly from that in the original Anderson (1991) paper, but the previously given specification was guided by John Anderson (personal communication; see also Pothos, 2007, and Pothos & Bailey, 2009, both of which employed the Rational Model as stated before).

The account of the category variability effect given by the Rational Model can be demonstrated with a straightforward example (our experimental investigation follows from this example). Consider two categories; the items of one category are more variable than the items of the other category. Suppose next that a novel exemplar halfway between the prototypes (means) of each of the two categories is presented; participants would be asked to classify this novel exemplar in one category or the other. Deriving predictions regarding the classification of the intermediate item from the Rational Model is straightforward. The model assumes that stimuli are presented sequentially and in the typical application of the model stimuli cannot be preassigned a classification. However, when there are two well-separated categories, the Rational Model will typically discover them, regardless of order of presentation of the stimuli. Therefore, in order to derive a prediction for the critical intermediate item, we can first present to the Rational Model the exemplars from the two categories, so that it can discover the two (intended) categories, and subsequently present the intermediate item. Then, the classification of the critical test instance can be examined by presenting it last.

As noted, the standard version of the Rational Model, as proposed by Anderson (1991) and applied by us, has one free parameter, c, the coupling parameter. Lower values of the coupling parameter make it less likely that dissimilar stimuli will be included in the same cluster and vice versa. The typical value of the coupling parameter is 0.5 and a typical range of exploring the coupling parameter, where relevant, would be between values close to 0 and 1. We explored predictions of the Rational Model regarding category variability across a range of values of the coupling parameter. The predictions of the Rational Model most relevant to our empirical situation correspond to values of the coupling parameter which lead to the intended classification. However, it is worth exploring the behaviour of the model across several values of the coupling parameter anyway, to examine the robustness of a

putative bias for classification into the more variable category. The input to the rational model was the stimulus coordinates for Experiment 1, as shown in Table 1a.

As expected, when the coupling parameter was very low all (or most) stimuli were assigned into separate categories ($c = 0.1$ and $c = 0.2$). Conversely, when the coupling parameter was high all items were included in a single category ($c = 1$). For $c = 0.3$ and 0.4, four separate groups were

TABLE 1
The test items and their dimensions, exactly as presented to participants in Experiment 1

Group name	Length of body in millimetres	Length of legs in millimetres
(a)		
1. S1	3	5
2. S1	3	5
3. S1	3	5
4. S1	3	5
5. S1	3	5
6. S1	3	5
7. S1	3	5
8. S1	3	5
9. S1	3	5
1. S2	8	6
2. S2	7	5
3. S2	6	4
4. S2	8	6
5. S2	7	5
6. S2	6	4
7. S2	8	6
8. S2	7	5
9. S2	6	4
Test item (S2)	5	5
(b)		
1. S1	8	6
2. S1	7	5
3. S1	6	4
4. S1	8	6
5. S1	7	5
6. S1	6	4
7. S1	8	6
8. S1	7	5
9. S1	6	4
1. S2	3	5
2. S2	3	5
3. S2	3	5
4. S2	3	5
5. S2	3	5
6. S2	3	5
7. S2	3	5
8. S2	3	5
9. S2	3	5
Test item (S1)	5	5

The test items were supposed to correspond to hypothetical spiders. Low values correspond to exemplars from the low variability category.

produced. Even though in these cases the halfway instance was classified into the low variability category, the low variability category was the largest one as well: the Rational Model has a bias favouring classification into larger categories (recall, *Prob*(category) α n_k, the number of items classified into the category so far). In any case, as noted, all these results are not relevant for our empirical situation, since the intended two-cluster classification was basically indicated to participants. For $c = 0.5$, 0.6, 0.7, 0.8, and 0.9, the Rational Model predicted the intended two-cluster partition of the stimuli. In all these cases, the critical test instance was classified into the more variable category. Thus, the Rational Model makes a robust prediction that a test instance halfway between two category proto-types should be assigned into the more variable category. This situation is unchanged if one considers a category of low variability (instead of zero variability) and a category of high variability (e.g., as in Experiment 3). Overall, as long as there are no large differences in category size, the Rational Model will predict that a P-halfway test instance will be classified into the more variable category.

It can be demonstrated intuitively why the Rational Model makes the prediction it does. Figure 2 shows two normal distributions with a straight line halfway between the means (proto-types) of the two distributions. This situation corresponds to the stimulus structures employed in our experiments (though in some of the experiments the low variability category structure has zero variability). The Rational Model works by examining how likely a new instance is, given membership into different candidate categories. A new instance is likely to be a member of a category if it is consistent with the distributional properties of the category. Clearly, in Figure 2, the instance is much more likely given the left (more variable) category, than the right (see also Hsu & Griffiths, 2010, and Stewart & Chater, 2002). This simple, but compelling, intuition is the basis for the classification predictions from the Rational Model (for a related illustration see Mareschal et al., 2002; for other examinations of the Rational Model, see Pothos, 2007, Pothos & Bailey, 2009, or Sanborn et al., 2010). In fact, the situation is entirely analogous to that in standard independent samples *t*-tests for comparing two means.

Note that we have chosen to examine the Rational Model in an idealised situation, without
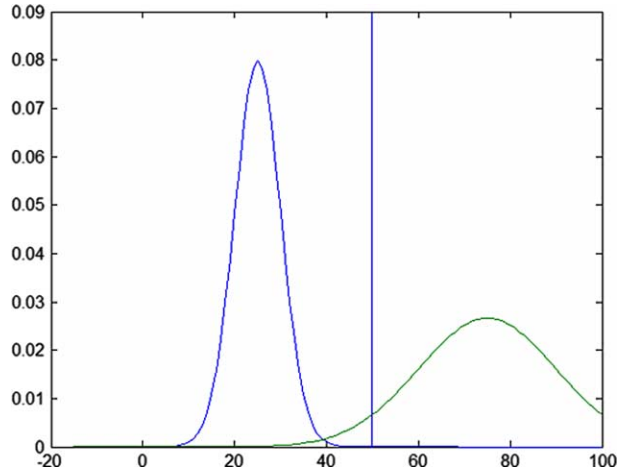
Figure 2. An illustration of why the Rational Model favours the more variable category when classifying a test instance halfway between the prototypes of the two categories. The test instance is more likely given the distribution of the more variable category.

reference to participant data. We chose this approach mainly for two reasons. First, because in this way we can explore the generic model biases regarding variability (that is, in the absence of complications that may arise from fits to noisy participant data). Second, because the standard version of the Rational Model allows relatively little flexibility in terms of particular fits anyway (the only parameter which can be manipulated is the coupling parameter). Overall, the main relevant insight regarding the psychology of categorisation from the Rational Model is that a bias to classify into the more variable category can arise if participants encode the distributional properties of the encountered instances and utilise them in a Bayesian scheme for the classification of new instances. It is worth stressing that the assumption that the categorisation process involves optimal Bayesian processing is an extremely powerful and ambitious one, and contrasts with other influential approaches for categorisation (e.g., prototype theory). Finally, the prediction regarding category variability from the Rational Model is robust, that is there do not seem to be circumstances that lead to the reversal of the prediction (unless there are major differences in the relative size of the categories).

EXPERIMENT 1

The first experiment was a basic test of the Rational Model prediction regarding category variability. In other words, we examined the prediction that a P-halfway test object will be classified as belonging to the high variability category. The procedure was designed to make any differences in category variability as salient as possible, as this appears to make it more likely that a critical exemplar (either halfway between the prototypes or halfway between the nearest neighbours) would be classified into the more variable category. This has been highlighted in at least three studies. For example, Smith and Sloman (1994), in an attempt to replicate Rips's (1989) finding, reported that participants were more likely to classify an NN-halfway critical instance in the high variability category if they could be shown (through their verbal protocols) to be aware that one category was indeed more variable. Stewart and Chater (2002, Exp. 1) observed higher classification of an NN-halfway test instance into the more variable category if participants were told, during training, that one of the two categories was more variable. Finally, Sakamoto et al. (2008) found that a random presentation of the same stimuli in a category made the category look more variable and so favoured it regarding the classification of an NN-halfway critical instance.

We adopted a novel approach to the problem of making category variability salient. Rather than presenting stimuli to participants, we showed them a table with the actual dimension values out of which the stimuli were meant to be constructed. In this way, participants could directly see the variability in each category (all participants were students in a psychology department, and so would have had some experience with processing tables of numbers of this sort). Such an approach avoids a number of potentially tricky issues associated with previous studies. First, there

would be no problem of having to establish whether the way participants represent the stimuli is the same way as the one assumed by the experimenter (e.g., see Cohen et al., 2001). The numbers presented to participants are the exact numbers used in the examinations of the Rational Model. Second, there are no issues arising from how well participants represent the differences between stimuli. Clearly, if participants fail to accurately perceive or represent stimulus differences, there is no point asking about whether category variability can affect classification (cf. Smith & Sloman, 1994). Third, it might be the case that an intuition about variability may be more difficult to ascertain with a small number of perceptual stimuli as opposed to a small set of numbers.

Regarding its other details, this experiment examined a P-halfway critical instance, rather than the nearest exemplars (Cohen et al., 2001; Stewart & Chater, 2002, Exp. 1; Sakamoto et al., 2008). Also, the category boundaries for the low and high variability categories, regardless of how the test instance was classified, were always simple linear ones (contrast with Cohen et al., 2001). In this way, we wanted to avoid any complications arising from category boundaries varying in complexity.

Methods

Participants. Participants were Swansea University undergraduate students, who volunteered to take part in the study. Twenty-two participants were presented with Table 1 (a or b).

Procedure and stimuli. Participants were presented with the flowing instructions:

A scientist in the Amazon discovered two types of spiders. The spiders are different in terms of length of bodies and length of legs. Nine spiders that belong to each of the two groups (S1 group and S2 group) are presented in one table as follows:

At that point, participants were shown either Table 1a or Table 1b, without the test item of course. The "a" and "b" versions of Table 1 correspond to whether the low variability category was called S1 or S2. We also counterbalanced whether the prototype of the low

variability category was "larger" than the prototype of the high variability category; in Table 1a and 1b the prototype of the low variability category was smaller, but we also had an alternative set of materials in which it was larger. Participants were subsequently asked the following questions:

In what group is there more variability between the spiders? 1. The S1 group or 2. The S2 group. A spider with 5 millimetre body and 5 millimetre legs is from: 1. The S1 group or 2. The S2 group.

As can be seen in Table 1, the low variability category consisted of repeating the same item several times, that is, it had zero variability. This was done as an additional manipulation to make the difference in variability between the two categories more salient. Moreover, the test instance was always exactly halfway between the prototypes of the two categories.

All materials were presented to participants as printed sheets of paper, and participants indicated their answers by circling the appropriate options. The experiment lasted approximately 5 minutes.

Results

All of the participants answered the first question correctly; the first question was meant to both increase the salience of category variability differences and to ensure that participants were attending to the task. As can be seen in Table 2, participants almost always classified the target item as belonging to the high variability group, consistent with the prediction of the Rational Model. This can be examined statistically in two ways. First, we can assume that factors *other* than category variability determine classification into the S1 or S2 category. In such a case, we would expect that classification of the test instance into the two categories will be the same, regardless of whether S1 is the more variable category or not. Thus, we can compare the frequency of classifications across the two conditions (S1 is the more variable category vs. S2 is the more variable category), with a chi-square test; this was found to be highly significant, $\chi^2(1) = 8.82$, $p < .05$. Second, we can examine the probability of classifying into the more variable category (regardless of whether this is S1 or S2), against a null hypothesis of chance classification into the more or less variable category. That is, this second

approach does not consider classification into S1 versus S2 as the main dependent variable, but rather classification into the more variable versus less variable category as the main dependent variable. A binomial distribution can be used to compute the probability of how likely it is to obtain k classifications into the more variable category, out of n, assuming an equal probability of classifying into the more and less variable categories. This probability was $p = .001$, so the null hypothesis that there is an equal classification probability into the less and more variable categories can be rejected. In other words, when S1 was the high variability category, the test instance tended to be classified to S1 and when S2 was the high variability category, the test instance tended to be classified to S2.

EXPERIMENT 2

In Experiment 1 we found that participants classified a test instance halfway between the prototypes of a low and a high variability category to the more variable category. However, no actual stimuli were presented. Such a manipulation was considered worth exploring in that it provided the closest possible match between the stimuli participants saw (the numbers in Table 1) and the input to Anderson's (1991) Rational Model. In Experiment 2, we adopted a more standard procedure, so as to examine the generality of the findings from Experiment 1. Half of the participants received only verbal descriptions of the categories (that is, participants would see Table 1a and 1b, as in Experiment 1), without any actual stimuli, and the other half received both the verbal descriptions and a set of actual objects (which were schematic spider-like forms).

TABLE 2
Classification of the test item in Experiment 1

	Test item was classified as S1	Test item was classified as S2
S1 is high variability	10	2
S2 is high variability	2	8

Here, and for all other tables reporting empirical results, cell numbers correspond to number of participants classifying the test item in different ways.

Methods

Participants. The population sample consisted of 109 participants, Swansea University undergraduates, who volunteered to take part in the study. Of these, 59 were presented with both the actual object and a verbal description and 54 participants were presented with only the verbal description.

Procedure and stimuli. Stimuli were constructed on the basis of the values on Table 1a and 1b which looked like spider-like schematic drawings (Figure 3; note that there was a second set of stimuli corresponding to the counterbalancing of the values in Table 1, as discussed for Experiment 1). The length of the legs and the length of the bodies corresponded to number of body parts and leg parts in the drawings. For example, in Figure 3 the spider on the left has four body segments (parts) and six leg segments. Additionally, together with each spider stimulus, the numbers of leg and body parts were presented verbally as well. For example, the spider on the right of Figure 3 was presented with the label "2 body parts and 4 leg parts". All of the stimuli in a group (S1 or S2) were presented on the same sheet of paper.

Participants received their instructions printed on an A4 sheet of paper. In the first condition, participants were presented with both the physical stimuli and the corresponding verbal labels. Once participants indicated that they had read and understood the instructions, they were given a second sheet (the stimuli in the S1 group) with the actual stimuli and their verbal descriptions and then with a third sheet (the stimuli in the S2 group; in all cases, the intended classifications were indicated). Note that the physical size of the stimuli was such that all the stimuli in each group could fit onto a single A4 sheet. The question regarding category variability and the classification of the item were asked in a final sheet, as has been the case in Experiment 1. The second condition was identical, but for the fact that participants saw only the verbal descriptions of the stimuli (this applies to the test stimulus as well).

Results

As can be seen in Table 3, participants who were presented with objects and verbal descriptions,
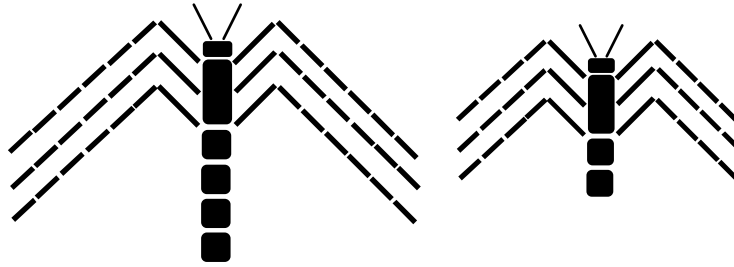
Figure 3.   An example of the stimuli that were presented to the participants in Experiment 2.

were more likely to classify the intermediate item in the high variability category, $\chi^2(1) = 13.665$, $p < .01$; binomial probability of chance classification was found to be .0002. Accordingly, when S1 was the high variability category, participants tended to classify the test item as belonging to that S1 group and when S2 was the high variability category, participants tended to classify the test item as belonging to the S2 group. Similar results were found for participants who were presented only with the linguistic descriptions, so that, as shown in Table 3, participants were more likely to classify the intermediate item in the high variability category, $\chi^2(1) = 21.47$, $p < .01$; binomial probability of chance classification < .00005.

In Table 4 we examine the effect of presenting the actual physical stimuli. It is clear that this manipulation had very little effect and that participants always preferred to classify the test item in the category that was more variable, $\chi^2(1) = 1.18$, $p > .1$. In other words, participants' classification preferences were the same regardless of whether training items were presented only verbally or with verbal descriptions combined

with actual pictures of the stimuli. The conclusion from both Experiments 1 and 2 is that when the variability of a category is made salient, participants classify a test stimulus halfway between the category prototypes in the more variable category, consistently with the predictions from the Rational Model.

EXPERIMENT 3

A key aspect in the procedure for Experiments 1 and 2 was that category variability was made salient in a number of ways (the stimuli were shown concurrently and were composed of segments that made stimulus differences obvious, participants were asked a question regarding category variability, and there was zero variability in one category). In Experiment 3, we investigated a design analogous to that of Experiments 1 and 2, but where category variability was less salient. This was achieved with the use of somewhat different stimuli (they were not composed of segments), dropping the question in the test phase relating to category variability, and using a sequential presentation procedure.

As discussed, previous, related research which employed a sequential stimulus presentation has led to mixed results. Cohen et al. (2001) and Sakamoto et al. (2008) observed that an

TABLE 3
Classification of the test item for participants who were presented with actual stimuli and linguistic descriptions or just the linguistic descriptions in Experiment 2

	Test item was classified as S1	Test item was classified as S2
Participants saw actual stimuli and linguistic descriptions		
S1 is high variability	24	12
S2 is high variability	4	19
Participants saw only linguistic descriptions		
S1 is high variability	18	3
S2 is high variability	7	26

TABLE 4
Classification of the test item, depending on whether participants saw pictures of the stimuli or not in Experiment 2

	Participants classifying to	
	The more variable category	The less variable category
Actual stimuli shown	43	16
Only linguistic labels	44	10

NN-halfway critical instance was more likely to be classified into the high variability category. By contrast, Stewart and Chater (2002) did not observe classification into the high variability category with a sequential presentation; such a result was obtained only when the stimuli were presented concurrently (as in our Experiment 2). Note, however, that all these studies employed a fairly extensive supervised categorisation procedure. With this experiment, we were interested in categories which would be highly intuitive, so that a single presentation of the stimuli would (or should) readily enable participants to understand them. This is a significant consideration in this context, because the Rational Model is a model of unsupervised categorisation, so that it cannot learn categories on the basis of feedback. Thus, the natural application of the Rational Model would be such that the intended categories are spontaneously recognised.

Finally, it is worth confirming the predictions of the Rational Model for the Experiment 3 stimuli (Table 5). For $c = 0.1$ all items were assigned to their own clusters. For $c = 1$ all items were assigned to the same group. For $c = 0.2, 0.3, 0.4, 0.5, 0.6$, and 0.7, between five and three clusters were produced and in all cases the P-halfway instance was individually classified. For $c = 0.8$ and 0.9 the intended two-cluster classification was produced and in both cases the P-halfway critical test instance was assigned to the more variable category. In other words, the Rational Model's prediction is unchanged: The model still predicts classification of a novel instance into the high variability category.

Methods

Participants. Participants were 66 Swansea University undergraduate students, who volunteered to take part in the study.

Stimuli. Stimuli were spider-like schematic drawings, analogous, but not identical, to the ones used in Experiment 2 (see Figure 4). The length of the legs and the length of the bodies were mapped to the values in Table 5 by assuming a Weber fraction of about 8%; such a Weber fraction is a reasonable estimate for our stimulus dimensions, since they effectively correspond to lengths (e.g., Morgan, 2005). Moreover, as in other experiments, we counterbalanced whether

the low variability category consisted of "small" or "large" stimuli. These considerations increase our confidence that the assumed representation was the actual psychological one, though we did not carry out a more detailed examination of this

TABLE 5
Length of legs and length of bodies for the stimuli in Experiment 3

Group name	Length of body	Length of legs
(a)		
1. A	0	1
2. A	1	2
3. A	1	0
4. A	2	1
5. A	1.5	1.5
6. A	1	1
7. A	1.5	1
8. A	1	1.5
SD	0.582482	0.582482
Mean	1.125	1.125
1. B	8	10
2. B	10	8
3. B	9	10
4. B	10	9
5. B	8	9.5
6. B	9.5	8
7. B	8.5	10
8. B	10	8.5
SD	0.876275	0.876275
Mean	9.125	9.125
Test item (B)	5.125	5.125
(b)		
1. B	8	10
2. B	10	8
3. B	9	10
4. B	10	9
5. B	8	9.5
6. B	9.5	8
7. B	8.5	10
8. B	10	8.5
SD	0.876275	0.876275
Mean	9.125	9.125
1. A	0	1
2. A	1	2
3. A	1	0
4. A	2	1
5. A	1.5	1.5
6. A	1	1
7. A	1.5	1
8. A	1	1.5
SD	0.582482	0.582482
Mean	1.125	1.125
Test item (B)	5.125	5.125

In this experiment participants only saw the actual stimuli. SD = standard deviation, which is a measure of variability of the members of the two categories. In this design, the test item is predicted to be classified to category B by the rational model.

issue (as, e.g., Cohen et al., 2001). The range of actual lengths for both legs and bodies was the same, and it was 0 mm (when there was no body or legs) to 76.1 mm. The smallest actual length (when it was 1) of a body or leg was 20 mm. In this experiment, each stimulus was printed individually onto a sheet of paper, which was cropped to be as large as the stimulus; stimuli were subsequently laminated. Finally, in this experiment, we created two new low/high variability categories, as seen in Table 5. Table 5 corresponds to stimuli such that the low variability category corresponds to low exemplars with generally low physical values and the high variability category corresponds to exemplars with high physical values. Another set of stimuli was created by reversing the correspondence between physical values and category variability; this was done so as to counterbalance whether the prototype of the low variability category was "larger" or "smaller" than the prototype of the high variability category. In all cases, the test item was a P-halfway item.

Procedure. Participants were first told that they would be presented with two groups of objects and that the objects differed in terms of the length of the bodies and the length of the legs; two examples of possible objects were then shown to illustrate the critical dimensions. They were also told that each of the two groups had eight members. The experimenter then started showing to participants the members of each group one by one. Each stimulus was printed on a laminated card and the stimulus order was the same as that on Table 5 (a or b; likewise for the alternative set of values, corresponding to the counterbalancing mentioned earlier). Once a stimulus had been shown, it was taken away. Finally, participants were told that they would have to decide to which group they wanted to assign a test stimulus they were about to see. Following the question, the test
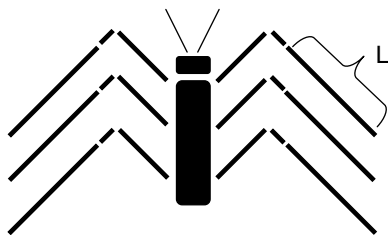
stimulus was presented and participants indicated their response orally to the experimenter.

Results

As can be seen in Table 6, participants tended to classify the target item as belonging to the *less* variable group, $\chi^2(1) = 3.956$, $p < .05$; binomial probability of chance classification = .014. In other words, when category S1 was the more variable one, participants tended to classify the test item as belonging to category S2, and likewise for category S2.

EXPERIMENT 4

Experiments 1, 2, and 3 provide conflicting intuitions regarding whether classification into the more variable category is always favoured or not. It appears that classification into the more variable category has to do with the extent to which differences in category variability are made salient or not. In Experiment 3, although the two categories did differ in relative variability, several manipulations were introduced to undermine the ability of participants to perceive the relative variability of the two categories. Specifically, in Experiment 3 the low variability category did not have zero variability, the question about variability (which was meant to function as a hint that the two categories differed in variability) was not included, and, finally, the stimuli were sequentially presented and were not made of individual segments (in Experiments 1 and 2 the stimuli were made of individual segments, so as to make any differences between different stimuli more salient). Under such conditions, there was a preference to classify the P-halfway instance into the less variable category.

We would like to understand in more detail the factors that contribute to the perception of a

Figure 4.   An example of the stimuli presented to participants in Experiment 3. The body is the black thick line in the middle and the legs are the black lines on the side.

TABLE 6
Classification of the test item in Experiment 3

	Test item was classified as S1	Test item was classified as S2
Category S1 is more variable	14	21
Category S2 is more variable	20	11

difference in category variability. In Experiment 4, we employed stimuli identical to those of Experiment 2 (spider-like images, composed of segments). The low variability category had zero variability. And the question about variability was included. But the stimuli in Experiment 4 were presented sequentially, instead of concurrently on a single sheet of paper (as in other experiments, participants were told the intended categories). Thus, there was a single difference between Experiment 2 and Experiment 4, namely that in the latter experiment the stimuli were presented sequentially. Sequential presentation is one of the factors highlighted in previous research as encouraging classification into the *less* variable category (Stewart & Chater, 2002). Indeed, in Experiment 3, which employed sequential presentation, there was a bias to classify in the less variable category. Thus, with Experiment 4 we try to isolate the importance of sequential versus concurrent presentation regarding the bias to classify into the more variable category.

Methods

We recruited 40 experimentally naïve participants from the undergraduate population of the Ben-Gurion University. The stimuli were identical to those of Experiment 2. The procedure was identical to that of Experiment 3, that is, each stimulus (with its intended category label) was individually presented to participants. As this experiment was run in Israel, all materials were translated into Hebrew.

Results

As can be seen in Table 7, participants tended to classify the target item as belonging to the *more* variable group, $\chi^2(1) = 12.91$, $p < .05$; binomial probability of chance classification $= .0002$. In other words, when category S1 was the more variable one, participants tended to classify the test item as belonging to category S1, and likewise for category S2. This result clearly shows that a change in presentation format, from concurrent to sequential, is in itself insufficient to overcome the bias to classify a P-halfway test item into the more variable category.

TABLE 7
Classification of the test item in Experiment 4

	Test item was classified as S1	Test item was classified as S2
S1 is high variability	18	2
S2 is high variability	7	13

EXPERIMENT 5

In Experiments 1, 2, and 3 we observed a bias for classification into the more and less variable categories respectively, but Experiments 1 and 2 differed from Experiment 3 in several ways. Experiment 4 rules out one of the possible relevant differences between Experiments 1, 2, and 3. Namely, Experiment 4 showed that even with a sequential presentation it is possible to obtain the bias for classification into the more variable category. However, Experiment 4 still differs from Experiment 3 in several ways. In Experiment 4, the less variable category still had zero variability (in Experiment 3 it did not), the stimuli were composed of segments (in Experiment 3 they did not), and the question about variability was included (in Experiment 3 it did not). Thus, there are several factors which might be contributing to overcome the bias for classification into the more variable category in Experiment 3.

Experiment 5 was designed to examine another one of these factors, specifically the question (hint) about variability. Previous researchers had noted that providing a hint regarding the difference in variability between the two categories increased classification of an NN-halfway test item into the more variable category (Stewart & Chater, 2002; see also Sakamoto et al., 2008; Smith & Sloman, 1994). In the present series of experiments, such a hint had the form of asking participants which category was more variable. Is this question about variability a critical factor in encouraging classification into the more variable category? Experiment 5 was identical to Experiment 4, but the question about category variability was dropped. (Recall, Experiment 4 was identical to Experiment 2, but for the fact that in Experiment 4 the stimuli were sequentially presented, whereas in Experiment 2 presentation was concurrent.)

Methods

We recruited 40 experimentally naïve participants from the undergraduate population of the Ben-Gurion University. The stimuli and procedure were identical to those of Experiment 4, but for the fact that the question about category variability was not included. As this experiment was run in Israel, all materials were translated into Hebrew.

Results

As can be seen in Table 8, participants tended to classify the target item as belonging to the *more* variable group but the results were not significant, $\chi^2(1) = 1.76$, $p > .1$; binomial probability of chance classification $= .057$. In other words, the results of Experiment 5 are exactly in between the results of Experiment 1, 2, and 4 and those of Experiment 3. Experiment 5 shows that not including the question about category variability is an important factor in whether participants classify into the more variable category or not.

GENERAL DISCUSSION

The empirical objective of this study was to reassess some of the conclusions regarding the impact of category variability on the classification of instances halfway between two categories. There are different ways in which a halfway instance can be defined. We chose to focus on test instances halfway between prototypes (P-halfway instances), because prototypes are undeniably important characteristics of a category. By contrast, some researchers have considered test instances halfway between the nearest neighbours of two categories (NN-halfway instances). We were reluctant to adopt this alternative approach. The nearest neighbours between two categories would be the least typical mem-

bers of the categories (in statistical terms, they might be outliers, though whether they are outliers or not would depend on the overall distributional characteristics of the relevant categories). Therefore, their importance in the classification of new instances is debatable. Even though standard exemplar models do not specifically weigh some category members more than others, whether this is indeed the case or not has yet to come under close empirical scrutiny.

Our experiments were designed to incorporate a number of factors that may impact on classification into the more variable category. These factors were: the form of the stimuli (whether they were made of segments or not; in the former case, stimulus differences would be more salient), stimulus presentation format (concurrent vs. sequential), the variability of the low variability category (if it is zero, then presumably the difference in variability between the low and the high variability categories is more obvious), and a question about which category was more variable (this was meant to encourage participants to appreciate that there was a difference in the variabilities of the two categories). Experiments 1 and 2 were designed to have in place several of the factors thought to encourage participants to appreciate the difference in the variability of the two categories. In these cases, we found that a test instance halfway between the two category prototypes tended to be classified to the more variable category. By contrast, in Experiment 3 most of these factors were eliminated, leading to a preference for classifying the halfway test instance as belonging to the low variability category.

Experiments 4 and 5 were designed to explore in more detail the conditions that make it more likely that a P-halfway test instance would be classified into the more variable category. Experiment 4 was identical to Experiment 2, but for the fact that the stimuli were presented sequentially instead of concurrently. We found that in Experiment 4 classification into the more variable category was still favoured. Experiment 5 was identical to Experiment 4 but for the fact that the question about variability was dropped. In that case the results showed no significant bias to classify in the low or high variability category. Thus, Experiments 4 and 5 provide a graded picture for how the pattern of results in Experiment 2 eventually lead to the pattern of results in Experiment 3 and highlight the relative importance of the corresponding experiment

TABLE 8
Classification of the test item in Experiment 5

	Test item was classified as S1	Test item was classified as S2
S1 is high variability	9	11
S2 is high variability	5	15

TABLE 9
A summary of the experiments and key results

	Stimuli	Hint	Low variability category	Presentation	Classification preference
Experiment 1	Table of values	Question of which category was more variable	Zero variability	Concurrent	High variability category
Experiment 2, Condition 1	Table of values	Question of which category was more variable	Zero variability	Concurrent	High variability category
Experiment 2, Condition 2	Table of values and actual stimuli	Question of which category was more variable; Size of stimuli directly indicated	Zero variability	Concurrent	High variability category
Experiment 3	Actual stimuli	No hint	Low variability	Sequential	Low variability category
Experiment 4	Table of values and actual stimuli	Question of which category was more variable; Size of stimuli directly indicated	Zero variability	Sequential	High variability category
Experiment 5	Table of values and actual stimuli	Size of stimuli directly indicated	Zero variability	Sequential	No significant result

characteristics. Table 9 provides a summary of the experiments and key results.

This pattern of results is partly consistent with previous relevant research. For example, as with previous researchers, we concluded that in general when the difference in variability between two categories was made salient, the test instance was more likely to be classified in the more variable category (Sakamoto et al., 2008; Smith & Sloman, 1994; Stewart & Chater, 2002). We also confirmed the importance of a hint regarding the difference in variability between the two categories (Stewart & Chater, 2002). But, we could not conclude that sequential versus concurrent presentation is a critical consideration in undermining the assumed bias to classify a half-way instance into the more variable category (Stewart & Chater, 2002).

It is worth providing a brief consideration of why we observed a bias for classification into the less variable category in Experiment 3, though a full examination of this issue is beyond the scope of our work. One possibility is that the processing of nonidentical stimuli when presented sequentially does not lead to memory traces that are stable enough and detailed enough, for participants to get an accurate sense of differences in category variability, or, indeed an accurate sense of the categories themselves. Under such circumstances, maybe the low variability category is favoured for the classification of the test instance simply because it is learned better (or, because for the low variability category participants find it easier to come up with a heuristic rule to represent it; cf. Ashby, Alfonso-Reese, Turken, & Waldron, 1998). That is, classification results in Experiment 3 may depend not on which category is more or less variable, but rather on which category had been better acquired.

Experiment 3 was actually the only experiment in which the low category variability had nonzero variability. Is this a critical factor in terms of the preference for the low variability category we observed in this experiment? We think this is unlikely. First, there is no theory as to why zero variability would be qualitatively different from low, but nonzero variability. Second, even given the fact that the stimuli in the low variability category were identical, it is arguable as to whether the psychological representations of the stimuli were exactly identical as well. Participants probably felt the stimuli were nearly the same, but to establish exact identity closer scrutiny would be required (note that this is our

speculation, we did not specifically assess this issue). Thus, our tentative suggestion is that a difference between zero and low variability is only one of the factors making a difference in the variability of categories salient, and does not impact on participant performance in another way.

To sum up, our results add support for the view that classification of a P-halfway test instance is more likely to be in a high variability category, as opposed to a low variability one, as long as the difference in category variability is salient. Our results complement the considerable related literature on NN-halfway test instances and consolidate some of the intuitions regarding when it is more likely for classification into the more variable category to occur.

We also explored the predictions from the Rational Model of categorisation (Anderson, 1991; Sanborn et al., 2010), which is the predominant Bayesian model of categorisation. We implemented the continuous version of the Rational Model, which clearly predicts that a P-halfway test item will be classified into the more variable category. It is worth noting that this prediction holds across a wide range of manipulations of its main parameter, the coupling parameter. Our results are mostly consistent with the Rational Model, since participants did indeed typically classify the P-halfway test instance into the more variable category. This suggests that perhaps the bias for classification of P-halfway instances into the more variable category could arise because of a representation of the distributional properties of category exemplars and an assessment of new instances in terms of such distributional properties. However, the Rational Model cannot explain, for example, the results in Experiment 3, whereby participants favoured classification into the less variable category. We can speculate as to which assumptions of the Rational Model could be relaxed to account for the Experiment 3 result. The most obvious culprit is the assumption regarding the input to the model. As already noted, perhaps the sequential presentation of the stimuli did not allow participants to develop the intended categories or appreciate the differences in the variability of categories (note that the Rational Model, and other related computational models, has no straightforward way to formalise this idea). However, as also noted, we cannot preclude the possibility that there are restrictions in the applicability of the psychological processes

assumed by the Rational Model and that, in some cases, alternative categorisation modes are engaged (cf. Ashby et al., 1998).

Nosofsky (1991) argued that certain versions of exemplar theory are consistent with the predictions of the Rational Model (for his argument he employed a precursor to the GCM). So, one would expect that predictions from the GCM would (typically) converge with predictions from the Rational Model. But, as noted, Cohen et al. (2001) found that a slightly reduced version of the GCM was not consistent with a bias for classification into the more variable category, at least for NN-halfway instances. What is the situation for P-halfway instances? Rather than fit the GCM to participant data, as has been the case for our examination of the Rational Model, we believe a more informative analysis of the model's properties would be provided by an idealised situation whereby the P-halfway instance (e.g., in Table 1a) is assigned with 100% probability to either the low or the high variability category—an examination of the model's sum of squares (SoS) error for the alternative classifications will reveal the model's bias regarding more vs. less variable categories.

We explored the full version of the GCM, subject only to the requirement that the category bias parameters be equal (cf. Cohen et al., 2001). The similarity function parameter was constrained to be between 1 and 2 (since the exponential and Gaussian similarity functions have been the only ones which have been supported in the literature) and the Minkowski power metric parameter was constrained to be between 1 and 2 as well (since the City block and Euclidean metrics are likewise the only ones that have been advocated as psychologically relevant). Attentional weights were allowed to vary freely between 0 and 1, subject to a constraint that they sum to 1. Finally, the upper limit of the sensitivity parameter was manipulated (with perfectly linear categories and perfect classification probabilities, the GCM will nearly always favour the max allowed value of the sensitivity parameter; Pothos & Bailey, 2009). For the Table 1a stimuli, we examined upper limits for the sensitivity parameter of 100, 50, 10, and 5, 1, and 0.5, and in all cases the SoS error was higher when the P-halfway instance was assigned to the low variability category than when it was assigned to the high variability category. Analogous results were obtained for the Table 3 stimuli.

These results show that for our stimulus sets the GCM favours classification of a P-halfway test instance into the more variable category. Intuitively, the GCM favours classification into a more variable category for P-halfway instances because of its exponentially decaying similarity function. A P-halfway test instance will be *closer* to the nearest neighbour from the high variability category, than the low variability one. Moreover, an exponentially decaying similarity function means that classification of the test instance will be more heavily influenced by near instances, whereas the influence of instances further away will rapidly (exponentially) diminish. The consistency of the Rational Model and the GCM in relation to P-halfway instances highlights the interrelatedness of the two models (Nosofsky, 1991).

In conclusion, we presented several new experiments regarding classification biases for a test instance halfway between the prototypes of two categories and examined the predictions from two important categorisation models. The emerging conclusion is that there is a bias for classification into the more variable category, although this bias disappears when it can be assumed that there is reduced opportunity for participants to perceive the difference in variability between the two categories. This conclusion is partly consistent with the Rational Model and the GCM.

REFERENCES

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

Anderson, J. R., & Fincham, J. M. (1996). Categorization and sensitivity to correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 259–277.

Ashby, G. F., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105,* 442–481.

Ashby, G. F., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154–179.

Bacon, F. (1898). *Novum organum.* London: George Bell Sons. (Original work published 1620).

Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory and Cognition, 29,* 1153–1164.

Busemeyer, J. R., Pothos, E. M., Franco, R., & Trueblood, J. (2011). A quantum theoretical explanation for probability judgment errors. *Psychological Review, 118,* 193–218.

Carey, S. (1985). *Conceptual change in childhood.* Cambridge, MA: Bradford Books.

Carnap, R. (1950). *Logical foundations of probability.* Chicago, IL: University of Chicago Press.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory and Cognition, 29,* 1165–1175.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 234–257.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14,* 357–364.

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology, 64,* 159–174.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory and Cognition, 33,* 289–302.

Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology, 43,* 243–273.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.

Horwich, P. (1982). *Probability and evidence.* Cambridge, UK: Cambridge University Press.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach.* Chicago, IL: Open Court.

Hsu, A. S., & Griffiths, T. L. (2010). Effects of generative and discriminative learning on use of category variability. In R. Camtrabone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society,* 242–247.

Lopez, A. (1995). The diversity principle in the testing of arguments. *Memory and Cognition, 23,* 374–382.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folk biological taxonomies and inductions. *Cognitive Psychology, 32,* 251–295.

Lopez, A., Gelman, S. A., Gutheil, G., & Smith, E. E. (1992). The development of category-based induction. *Child Development, 63,* 1070–1090.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309–332.

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science, 26,* 377–389.

Morgan, M. J. (2005). The visual computation of 2-D area by human observers. *Vision Research, 45,* 2564–2570.

Nagel, E. (1939). *Principles of the theory of probability.* Chicago, IL: University of Chicago Press.

Nosofsky, R. M. (1986). Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700–708.

Nosofsky, R. M. (1991). Relation between the rational model and the context model of categorization. *Psychological Science, 2*, 416–421.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning.* Oxford, UK: Oxford University Press.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97*, 185–200.

Peterson, M. J., Meagher, R. B., Jr., Chait, H., & Gillie, S. (1973). The abstraction and generalization of dot patterns. *Cognitive Psychology, 4*, 378–398.

Posner, M. I., Goldsmith, R., & Welton, K. E., Jr. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73*, 28–38.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*, 353–363.

Pothos, E. M. (2007). Occam and Bayes in predicting category intuitiveness. *Artificial Intelligence Review, 28*, 257–274.

Pothos, E. M., & Bailey, T. M. (2009). Predicting category intuitiveness with the Rational Model, the Simplicity Model, and the Generalized Context Model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1062–1080.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303–343.

Pothos, E. M., Edwards, D. J., & Perlman, A. (2011). Supervised vs. unsupervised categorization: Two sides of the same coin? *Quarterly Journal of Experimental Psychology, 64*, 1692–1713.

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*, 463–475.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York, NY: Cambridge University Press.

Ruts, W., Storms, G., & Hampton, J. (2004). Linear separability in superordinate natural language concepts. *Memory and Cognition, 32*, 83–95.

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.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory and Cognition, 22*, 377–386.

Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 893–907.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*, 309–318.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science, 331*, 1279–1285.

Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin and Review, 15*, 732–749.

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development, 56*, 1574–1583.