As time goes by: The role of variability in category belief revision

Lauren Kearney (lkearney@psy.unsw.edu.au)
Brett K. Hayes (B.Hayes@unsw.edu.au)
School of Psychology, University of New South Wales
Sydney, 2052, Australia

Abstract

Two studies examined the effect of category variability on revision of stereotypical and non-stereotypical beliefs. In Experiment 1, participants learned the characteristics of various groups via exposure to numerical distributions of category attributes. These distributions had the same mean values, but either high or low levels of variability around the mean. The categories were either stereotyped categories or non-stereotyped categories. After learning category attributes, participants were exposed to disconfirming exemplars (drawn from a sample with higher or lower means). Participants were then asked to re-estimate the central tendency of the category and to rate the likelihood that the category had changed. Beliefs about categories with low variability were more influenced by disconfirming information than beliefs about highly variable categories. This relationship was the same across category domains. This same general pattern was found in Experiment 2 where participants had to request information about test instances before deciding whether category values had changed. The results are consistent with Bayesian models of belief revision.

Keywords: Categorization, Belief revision, Bayesian statistics

Introduction

Imagine that you were a climate change researcher and you discovered that the average number of cyclones in a particular location had increased by 30 per cent over the past three years. You then have to decide whether this represents good evidence for change in the underlying weather patterns. To answer this question one needs to consider not only the mean number of cyclones in the past, but also the past variability of cyclone activity. If the mean number of cyclones in the area had shown little variability prior to the recent increase, then you may be more convinced that some underlying change had occurred than if you knew that the past pattern of cyclone activity was highly variable.

This sort of problem can be construed as one of category belief revision. After encountering instances with unexpected properties can we conclude that the typical properties (e.g., annual frequency) of a category (e.g., cyclones) have changed? It seems evident that in order to remain useful, knowledge and beliefs about category properties need to reflect changes in underlying category structure (Garcia-Marques & Mackie, 1999). The challenge then for perceivers is to determine when meaningful change has actually occurred. As highlighted in the earlier example, one factor that seems important in determining when it is appropriate to revise category beliefs is the variability of the category exemplars that form the basis for the belief.

Although some work has examined the influence of category variability in related areas such as category learning (Stewart & Chater, 2002), category discrimination (Hahn, Bailey & Elvin, 2005), and induction (Nisbett, Jepson, Kunda and Krantz, 1993), relatively few studies have directly examined the impact of category variability on belief revision. In one of the few studies to examine this issue, Rehder and Hastie (1996) presented participants with category attributes. Each attribute was represented by a numerical value (e.g., a politeness score). Half the participants saw categories with highly variable attribute distributions while the other half saw distributions with low variability. Critically, the high and low variable distributions had the same mean; hence participants were expected to acquire similar beliefs about the central tendency of category attributes, but have different expectations about the level of variability around this value. When participants were subsequently exposed to novel exemplars with highly discrepant attribute values (higher or lower), participants who had initially seen the low variable distribution changed their central tendency beliefs more than did those who had initially seen the high variable distribution.

In addition to the experimental evidence, Rehder (1996) proposed a normative model outlining how people might use variability information. If the task is cast as one of Bayesian inference, whereby observation of a set of category exemplars allows one to calculate a central tendency estimate, then the variability of the original information should be of paramount importance. When one is presented with novel information, the calculation of the most likely posterior mean should be made on the basis of the prior mean, the variability of the initial distribution of exemplars, and the new single observation. As the standard deviation of the initial distribution increases, the weight of each piece of disconfirming information decreases and vice versa.

Although the Rehder and Hastie (1996) results support a Bayesian model of belief revision, some evidence suggests that people do not always respond to disconfirming information in a normative manner. One obvious domain where disconfirming information often fails to change existing beliefs is in social stereotyping (see Hilton & von Hippel, 1996 for a review). Moreover, there is evidence to suggest that, in the social domain, category variability may play a very different role in mediating belief change than the

one outlined by Rehder and Hastie (1996). Much of this work is concerned on a phenomenon known as the outgroup homogeneity effect (OHE). This refers to the tendency of members to perceive their own group (ingroup) as more heterogeneous and diverse as compared to members of other groups (outgroups) (e.g., Linville, Salovey & Fischer, 1989). Contrary to the normative approach, much existing evidence suggests that seeing outgroups as homogeneous (i.e. low in variability) leads to *greater resistance to change* following disconfirming evidence, (e.g., Hewstone, Johnston & Aird, 1992, Hewstone & Hamberger, 2000). Hewstone et al. (1992), for example, found that after viewing counter-stereotypical exemplars, there was more change in trait ratings for social groups who were perceived as having highly variable attributes (teachers) than for groups with low perceived variability (accountants). Such results imply that there may be some fundamental difference in the way that variability moderates belief revision in social and non-social domains.

The main aim of the current work was to examine this discrepancy in the role played by category variability in belief revision in non-social and social domains. Previously, a direct comparison of the role of variability in non-social and social categories has been difficult because each has been studied using very different methods. Studies in each domain have operationalized category variability in different ways, used different kinds disconfirming data and measures of belief change (cf. Hewstone et al. 1992; Rehder & Hastie, 1996). It is possible that these differences in stimulus structure and presentation, rather than domain per se, may have led to the discrepant findings with regard to role of variability in belief revision.

In the current studies, therefore, we used a common learning paradigm to manipulate variability in non-stereotypic (object) and stereotypic (social group) categories, and to test for revision of category beliefs following disconfirming data. Following Rehder and Hastie (1996) we created a set of object categories where the central tendency of category attributes was held constant but attribute variability differed. A corresponding set of stereotyped categories was created with the same central tendency but where there were pre-existing differences in participants' beliefs about the variability of the groups around that central value. These categories were designed to mimic the outgroup homogeneity effect such that there was greater attribute variability (but not central tendency) for participant ingroups ("City Dwellers", "Youths") than for corresponding participant outgroups ("Country Dwellers", "Seniors"). After learning these categories all participants were presented with three disconfirming instances (with attribute values above or below the training mean). Revision of category beliefs following disconfirmation was examined by asking people to re-estimate the category mean and rate the likelihood that this mean had changed.

If people use information about category variability in fundamentally different ways in the stereotyped and non-stereotyped domains then this should be reflected in the test data (with category heterogeneity at training leading to less belief revision in the non-stereotyped domain and more in the stereotyped domain). If, however, people use variability information in a normative way in both domains then we should see parallel patterns of belief revision in the stereotyped and non-stereotyped domains.

Experiment 1

Method

Participants
Forty five male and female Caucasian University of New South Wales undergraduate students (M_{age} = 19.3 years) participated in return for course credit.

Design and Materials
During the training phase, people learned the attributes of three non-stereotyped and three stereotyped categories. Each training category consisted of 32 instances with a single numerical attribute. The variability of these attribute values was manipulated between subjects by varying the standard deviation of category attributes around a common mean (see Table 1).

Table 1: Training and test distributions

Case		Training			Test	
	M	Low *SD*	High *SD*	Low *M*	High *M*	*SD*
A	165	8	24	141	189	8
B	64	6	18	46	82	6
C	19.8	4	12	7.8	31.8	4
D	5.7	1	3	2.7	8.7	1
E	345	11	33	345	345	11
F	9.1	2	6	9	9	2

Table 2: Categories and attributes

Case	Category	Attribute
A	New Zealanders	Light Sensitivity
B	Fiction Novels	Length
C	Seniors/Youths	Attitudes towards drugs
D	Asians/Caucasians	Politeness
E	College Students	Oxydilic Mineral Levels
F	Medicine/Psychology Students	Study hours per week

High variability categories had a standard deviation that was three times that of the low variability categories. For stereotyped categories low variability cases always involved participant outgroups (i.e. Asians and Seniors) and high variability cases always involved participant ingroups (i.e. Caucasians and Youths). Furthermore, the stereotyped categories were pilot tested to check that participants'

beliefs about the variability of in-groups and out-groups groups were in line with the outgroup homogeneity effect. Participants were asked to indicate the possible range of attribute scores for their ingroup and a corresponding outgroup. For the stereotyped categories utilized in this study, participants gave reliably wider range estimates (but not central tendency estimates) for their ingroups as compared to the corresponding outgroups, as predicted by the OHE (e.g., Linville, et al. 1989).

The non-stereotyped categories were drawn from Rehder and Hastie (1996). For the full list of the categories and attributes used see Table 2.

At test participants in each condition were shown three new instances (disconfirmers) from each of the training categories. For four of the categories (A-D), the test phase values were substantially different from the mean training values (see Table 1). These were designated "change" cases. For half the participants the values of test instances were above the training mean. For the remainder they were below the training mean. The other two categories (E-F) were "no change" cases in that test instances were drawn from the same distribution as the corresponding training instances.

Procedure

Participants were randomly allocated to a low or high category variability condition. They were told that they were to learn about the characteristics of a number of objects/groups that were measured on a critical attribute in the year 1995. Following a warm-up example, each participant was trained and tested on six categories (four change and two no change). The order of presentation of the six categories was randomly determined subject to the constraint that a change case was always presented first. In the training phase for each category 32 instances containing a single numerical value were presented on screen one at a time for 2s each, with a blank screen inter-stimulus interval of .3s. After viewing these values, participants were prompted to estimate the mean of the attribute scores and the maximum and minimum values observed. Immediately after completing these judgments, the test phase began. Participants were told that a more recent study of the same target category had been carried out (circa 2005) and that they would be presented with some of the results of this more recent study. Prior to the presentation of the test instances, participants were prompted to estimate the mean attribute value of the 2005 category. They were also asked to rate the likelihood that the mean value of the category for the test items had changed on an eleven-point scale (1= change since 1995 was very unlikely; 10 = change since 1995 was very likely). These judgments were referred to as the 0 disconfirmer phase, as participants had seen no new information. The three test instances/disconfirmers (from the 2005 data set) were then presented one at a time in the same manner as the training instances. After each disconfirming instance participants gave mean estimates and ratings of change likelihood. After completing the test phase, the training phase for the next category commenced. The procedure took approximately 45 minutes to complete.

Results and Discussion

To make the results of the different categories comparable, estimates of the category means and upper and lower bound estimates were converted into z scores using the training mean for the scenario and the standard deviation from the high variability version of the original distributions. Judgments of the range of training and test distributions were calculated by subtracting the z score of the lower bound estimates from the z score of the upper bound estimates. For the "change" categories the mean and range estimates for two stereotyped categories were collapsed to give a singe set of results for the stereotyped condition. The same was done for the two non-stereotyped categories. Change-likelihood scores for all four change cases were also collapsed into two sets of scores, one stereotyped and one non- stereotyped case.

Training Phase

Estimates of category means after training were generally accurate (A perfect estimate would have a z score of zero). In the high variability condition the mean estimates in z score units were .12 and .00 for the social and non social cases respectively. In the low variability condition the corresponding mean estimates were -.02 and -.02.

Estimates of the range of attribute values were also sensitive to the manipulation of the variability of the distribution. In the low variability condition, range estimates were 1.58 and 1.29 (in z score units) for the stereotyped and non-stereotyped cases respectively. The corresponding range estimates in the high variability condition (3.54 and 4.27 for social and non social), were significantly higher, $F (1, 43) = 1326.75$, $p < .001$. These results show that participants in both variability conditions accurately encoded the distributional characteristics of category attributes

Test Phase

Separate test analyses were carried out for change and no change categories. For "change" cases the mean value of the disconfirming instances was always either higher than the training mean (high disconfirmer) or lower than the training mean (low disconfirmer). To make these cases comparable, the z scores for participants in the low disconfirmer condition were multiplied by negative one. The high and low disconfirmer results were collapsed for each of the variability and domain conditions.

Test phase mean estimates (see Figure 1) were entered into 2 (variability condition) x (2) (stereotyped vs. non stereotyped) x (4) (number of disconfirmers: 0, 1, 2 or 3) ANOVA with repeated measures on the second and third factors. Estimates of the new mean were clearly influenced by the amount of disconfirming information, with a significant linear effect of number of disconfirmers on mean estimates $F (1, 41) = 51.01$, $p<.001$. Critically, there was also a significant interaction between the variability of the training category and linear trend for disconfirming instances, $F (1, 41) = 6.16$, $p < .05$. In line with the findings of Rehder and Hastie (1996), this linear relationship was stronger in the low variability condition than in the high

variability condition. As the number of disconfirmers increased, mean estimates in the low variability condition showed increasingly large deviations from the training mean. Critically, the effect of variability on belief revision was the same across both domains (stereotyped vs. non stereotyped), F (1, 41) = .490, p = .488. Contrary to previous work on the OHE (e.g., Hewstone et al., 1992), mean estimates for outgroups (low variable categories) were revised more in the face of disconfirming information than mean estimates for ingroups (highly variable categories).
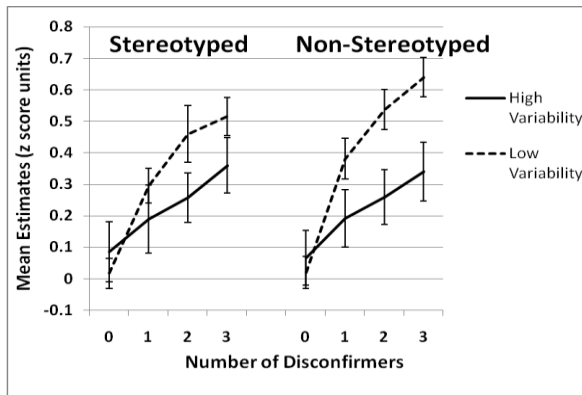

Figure 1. Test phase estimates of category means as a function of disconfirmers, variability and domain.

Similar results were obtained for change-likelihood scores (see Figure 2). There was a significant linear effect of number of disconfirmers on change-likelihood, F (1, 41) = 13.65, p < .001, with the likelihood of change increasing with the presentation of successive disconfirmers. Again, this effect was moderated by a significant interaction between linear trend and variability, F (1, 41) = 13.23, p < .001, Figure 2 shows that with successive disconfirmers, people in the low variability condition became increasingly likely to judge that the typical values of the category had changed. Those in the high variability condition however showed little change in their likelihood ratings across disconfirmers.
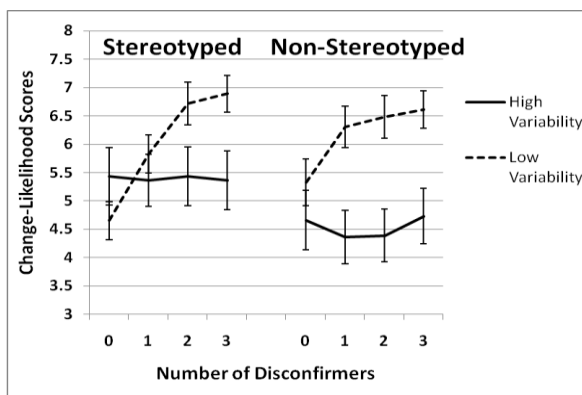

Figure 2. Change-likelihood scores as a function of disconfirmers, variability and domain

Analysis of the "no change" categories confirmed that neither variability, number of disconfirmers nor domain affected mean estimates or change likelihood ratings for these cases (all F's < 2)

Overall, the results suggest that variability in category attributes plays a similar role in belief revision for stereotyped and non-stereotyped domains. In both domains exposure to disconfirming information was more likely to lead to a perception of category change when categories were homogenous than when they were heterogeneous. These results suggest that in an environment where exposure to category members is externally controlled and people are not given the option of selectively sampling or discounting certain exemplars, variability information is used in a normative way when considering disconfirming information. This seems to be the case regardless of the category domain.

For this finding to be generalisable however, it is important to consider that exposing people to discrepant information may involve different processes than allowing people to actively seek it. For example Johnston (1996) compared the impact of forced exposure to counter-stereotypic information with free (participant controlled) exposure. Results indicated that only those in the forced condition showed moderation of their stereotyped judgments. This suggests that giving people the option of sampling new data to test whether or not a category has changed may result in the use of additional information processing strategies. These could include re-interpretation or discounting of the information (Johnston, 1996), which seems more likely to occur when there is a motivation to maintain the existing belief (e.g., in the case of stereotypes). Given that there is some evidence that higher levels of variability are associated with weaker stereotype effects and less confidence in group stereotypes (Judd, Ryan & Park, 1991) then one might expect that discounting or re-interpretation of discrepant information might be more likely to occur for homogeneous group stereotypes (i.e. those relating to outgroups). In this way, people may require more evidence to be convinced to change their outgroup stereotypes as compared to stereotypes about their ingroups.

Experiment 2 addressed this issue by allowing participants to actively control the amount of disconfirming information to which they were exposed.

Experiment 2

Method
Participants
40 University of New South Wales undergraduates (M_{age} = 19.3 years) participated in return for course credit.
Design and Procedure
The training phase stimuli and procedure were identical to Experiment 1. In the test phase, however, participants were asked to imagine themselves as field researchers and told that it was their task to sample from the new (circa 2005) values for each training category until they had determined

whether or not the attributes of the test categories differed from those encountered at training. On each test trial participants were asked whether they would like to see another observation from the test category. If they answered "yes" then a new value was presented on the screen for 4s. They were instructed to continue this process until they were "more than 75% certain that the 2005 category was either the same or different from the 1995 category" or until a total of 32 observations for a given category had been obtained. The general instructions emphasised that they should not draw more observations than was necessary to make their judgment about change as this would increase the costs of their research. The 32 test items for each category were generated from the test distributions given in Table 1. When a new observation was required the computer program sampled randomly without replacement from this set of instances. As in Experiment 1, four of the test categories were change cases where the mean of the test distribution was three standard deviations above or below the training mean. Two were no-change cases where the means were the same as the training distributions. After the participant had finished sampling they were asked to estimate the mean of test category and rate the likelihood that this mean was different from the training mean.

Results and Discussion
Training Phase
Once again z score estimates of the mean values of the six training categories were close to zero ($M's$ = 0.009, -0.023, 0.011, 0.126 for non-stereotype low variability, non-stereotype high variability, stereotype low variability, and stereotype high variability respectively). Again, the observed variability of category values was reflected in higher range z-score estimates being given to high variability (M= 3.76), than low variability categories (M = 1.37), $F (1, 38) = 550.69$, $p < .001$. In this experiment the no-change cases were treated as fillers and were not analyzed.

Test Phase
The key prediction involved the number of test instances sampled before a decision about category change was made. For change categories a 2 (variability) x 2(domain) analysis of variance found a significant effect of variability. Those trained with high variability categories requested more observations at test to make a decision about category change ($M = 7.61$) than those trained with low variability categories ($M = 5.94$), $F (1, 38) = 4.38$, $p < .05$. There were no significant differences between the number of instances searched for non-stereotypical ($M = 6.65$) and stereotypical categories ($M = 6.9$), and no interactions between domain and variability (all $F's < 1.0$).

General Discussion
These studies were concerned with whether people respond to change in the characteristics of categories and what role the variability of the original category exemplars might play in this process. Previous work on this issue has yielded

inconsistent results. Some (e.g., Rehder & Hastie, 1996) have demonstrated that when faced with novel and contradictory category information, low attribute variability can be conducive to changes in beliefs about category characteristics, whilst others have shown the opposite effect in the realm of stereotype change (Hewstone et al. 1992; Hewstone & Hamberger, 2000).

Consistent with the results of Rehder and Hastie (1996), both our studies showed that people are sensitive to changes in the central tendency of both stereotyped and non-stereotyped categories. Furthermore, variability appears to moderate responses to these changes, with beliefs based on less variable category information being more susceptible to revision than beliefs based on highly variable categories. In Experiment 1, two indexes of belief change were moderated by variability. As the amount of discrepant information increased, mean estimates in the low variability condition showed increasingly large deviations from the training mean. This pattern was mirrored by increasing ratings of the likelihood that the category central tendency had changed. Critically, the effect of category variability was observed regardless of whether the categories in question were stereotyped (ingroups and outgroups) or socially neutral. This was despite the fact that participants had pre-existing stereotypical beliefs about the relative variability of the ingroups and outgroups described in the experiment. This pattern of updating and sensitivity to variability information is consistent with a Bayesian model of belief revision (Rehder & Hastie, 1996). Experiment 2 showed that the observed relationship between variability and belief change holds even when people control the amount of disconfirming information they are exposed to.

Our findings provide converging evidence that experienced variability has a similar effect on belief revision in stereotyped and non-stereotyped categories. Stereotypical beliefs about groups with a homogeneous distribution of attributes seem to be more prone to revision in the face of disconfirming evidence than stereotypical beliefs about groups with a heterogeneous distribution of attributes. Category domain does not seem to materially affect the process of mean estimate revision.

These data support the notion that stereotype formation and maintenance may be a more specific instance of the general process of categorization (cf. Rothbart, 1981). The question still remains however why some previous studies on stereotyping have found more change for heterogeneous groups as compared to homogeneous groups (e.g. Hewstone et al. 1992, Hewstone and Hamberger, 2000). There are several possible explanations for this discrepancy. On a methodological note, in both the Hewstone et al. (1992) and Hewstone and Hamberger (2000) studies, the manipulation of variability was indirect and the disconfirming information was in a verbal format (descriptions of anti-stereotypic behaviours). The differences in the way the two studies operationalised variability and disconfirming information may be quite important for a number of reasons. Firstly, it has been demonstrated that people are able to

derive and utilize variability information in a relatively normative fashion when the information is presented in numerical distributions (Kareev, Amon & Horwitz-Zeliger, 2002) whereas it is less clear that people can derive statistically meaningful group variability information from implicit group stereotypes.

With regards to the disconfirming information, numerical disconfirming instances may be processed differently to disconfirming instances in the form of behavioral descriptions. Evidence from Nisbett, et al. (1983) suggests people find it difficult to reason statistically about social variables, especially when they do not have clear units of measurement. These difficulties may have been exacerbated by the trait statements utilised by Hewstone et al. (1992). Future work should aim to explore the role that mode of data presentation may have on the relationship between variability and belief change.

On a theoretical note, perceptions of variability in social domains may be laden with additional factors like causal theories about how group membership is related to group characteristics like the central tendency and variability of social behaviours (Yzerbyt, Rocher & Schadron, 1997)). For example, one might believe that Japanese people are uniformly polite because their culture is more collectivist and values conformity. When equipped with these causal theories, people may be less likely to relinquish their beliefs in the face of disconfirming information (Anderson, Lepper & Ross, 1980; Fugelsang, Stein, Green & Dunbar, 2004). Examining the influence of causal beliefs on the relationship between variability and belief change would be a productive avenue for future research.

Acknowledgments

References

Anderson, C.A., Lepper, M.R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*, 1037-1049.

Fugelsang, J.A., Stein, C.B., Green, A.E., & Dunbar, K.N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and cognitive laboratory. *Canadian Journal of Experimental Psychology, 58*, 86-95.

Garcia-Marques, L., & Mackie, D.M. (1999). The impact of stereotype-incongruent information on perceived group variability and stereotype change. *Journal of Personality and Social Psychology, 77*, 979-990.

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

Hewstone, M., & Hamberger, J. (2000). Perceived variability and stereotype change. *Journal of Experimental Social Psychology, 36*, 103-124.

Hewstone, M., Johnston, L., & Aird, P. (1992). Cognitive models of stereotype change: (2) Perceptions of homogeneous and heterogeneous groups. *European Journal of Social Psychology, 22*, 235-249.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47*, 237-271.

Johnston, L. (1996). Resisting change: Information seeking and stereotype change. *European Journal of Social Psychology, 26*, 799-825.

Judd, C. M., Ryan, C. S., & Park, B. (1991). Accuracy in the judgment of in-group and out-group variability. *Journal of Personality and Social Psychology, 61*, 366-379.

Kareev, Y, Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General, 2*, 287-297.

Linville, P. W., Salovey, P., & Fischer, G. W. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57*, 165-188.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339-363.

Rehder, B. (1996). A Bayesian analysis of the relationship between category variability and belief revision. *Unpublished manuscript.*

Rehder, B., & Hastie, R. (1996). The moderating influence of variability on belief revision. *Psychonomic Bulletin and Review, 3*, 499-503.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and ingroup behaviour*. Hillsdale, New Jersey: Erlbaum.

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

Yzerbyt, V. Y., Rocher, S. J., Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. Oakes, N. Ellemers A. Haslam (Eds.), *The social psychology of stereotyping and group life*. Oxford: Basil Blackwell.