Original citation:
Spicer, Jake and Sanborn, Adam N. (2017) A rational approach to stereotype change. In: CogSci 2017 : 39th Annual Meeting of the Cognitive Science Society, London, UK, 26–29 Jul 2017. Published in: Proceedings of the 39th Annual Conference of the Cognitive Science Society (In Press)

Permanent WRAP URL:
http://wrap.warwick.ac.uk/90820

A Rational Approach to Stereotype Change

Jake Spicer (J.Spicer@warwick.ac.uk)
Department of Psychology, Humanities Building,
University of Warwick, Coventry, UK

Adam Sanborn (A.N.Sanborn@warwick.ac.uk)
Department of Psychology, Humanities Building,
University of Warwick, Coventry, UK

Abstract

Existing theories of stereotype change have often made use of categorisation principles in order to provide qualitative explanations for both the revision and maintenance of stereotypical beliefs. The present paper examines the quantitative methods underlying these explanations, contrasting both rational and heuristic models of stereotype change using participant data and model fits. In a comparison of three models each simulating existing descriptions of stereotype change, both empirical data and model fits suggest that stereotypes are updated using rational categorisation processes. This presents stereotype use as a more rational behaviour than may commonly be assumed, and provides new avenues of encouraging stereotype change according to rational principles.

Keywords: Stereotypes; Categorisation; Rational Behaviour

Introduction

Stereotypes have often been found to be resistant to change, with beliefs and expectations regarding a group often persisting even when faced with directly contradictory information (Hilton & von Hippel, 1996). This presents a problem when trying to combat stereotypes underlying prejudice or discrimination through out-group exposure as has often been suggested by theories such as the Contact Hypothesis (Allport, 1954), as there is no assurance that simply demonstrating the inaccuracy of these beliefs will be effective in encouraging revision. It is therefore necessary to examine the processes by which stereotypes are updated with experience, and, in cases of stereotype persistence, determine how counter-stereotypical information can be disregarded in order to develop better methods to encourage change.

Past research into this field has offered three possible processes of stereotype revision: *book-keeping*, in which the stereotype is slowly adjusted with each relevant observation; *conversion*, in which the stereotype can undergo sudden and drastic changes in response to particularly notable contradictory exemplars; and *subtyping*, in which counter-stereotypical evidence is isolated from the rest of the category in a distinct subgroup, ignored when making category judgements. This presents three potential explanations for stereotype persistence: stereotype-incongruent exemplars may be noted via book-keeping but remain out-weighed by prior stereotypical beliefs; these exemplars may not have been sufficiently significant to evoke change via conversion; or these exemplars may have been excluded entirely via subtyping.

This distinction was examined by Weber and Crocker (1983) by manipulating the presentation format of counter-stereotypical evidence in summaries of lawyers: equal amounts of stereotype-incongruent evidence were either concentrated into only a few exemplars, or dispersed across many exemplars. This generates three competing expectations between the three models: conversion suggests that these concentrated exemplars would act as extreme disconfirmers, encouraging significant revision to the stereotype. Conversely, subtyping would suggest that concentrating incongruent evidence should make it easier to isolate, thereby preserving existing stereotypical beliefs. Book-keeping, meanwhile, focuses only on the amount of data rather than the presentation format, and so suggests no difference between these conditions. Measures of the strength of stereotypical beliefs following exposure to these exemplars were found to be stronger in the concentrated condition, supporting the subtyping model, an effect that has since been replicated in a number of studies (Bott & Murphy, 2007; Johnston & Hewstone, 1992).

This depicts stereotype persistence as an issue of categorisation, occurring where counter-stereotypical group members are placed in a distinct subgroup rather than integrated into existing structures. The mechanisms underlying stereotype revision could then be well described by existing models of categorisation, particularly those which perform a similar process of partitioning a category into lower-order subgroups. One key example of such a model is the Rational Model of Categorisation (RMC) developed by Anderson (1991), which organises a category into 'clusters' of exemplars based on similarities in observed features. The organisation of these clusters then determines the impact of stored data on subsequent judgements, with larger clusters tending to have more influence on expectations of given traits appearing in the category.

The subtyping effect could therefore be seen as the result of standard categorisation processes creating partitions of the category based on observed data patterns which determine the impact of incongruent information on later judgements: isolating this information leads to subtyping, diminishing its impact, while integration of congruent and incongruent information leads to book-keeping, and so greater stereotype revision. If so, subtyping may not be the result of a bias towards ignoring counter-stereotypical data in order to preserve stereotypical beliefs, but a rational incorporation of all available data under certain data patterns which happens to mitigate the influence of stereotype-incongruent information. As such, subtyping could be considered a more rational process than it may initially seem, and so could be fought using similarly rational mechanisms to encourage stereotype change.

The present study therefore presents a rational approach to stereotype use; in the following paper, we develop several candidate models to approximate the existing depictions of stereotype revision, contrast the predictions of these models with participant data to assess their accuracy, and use these findings to offer some insight into the process of stereotype change.

Model Details

We began by developing an edited version of the RMC in order to examine the categorical explanation proposed for the subtyping effect. This made use of the standard version of the RMC using discrete exemplar dimensions as defined by Anderson (1991), chosen for its reasonable level of simplicity and ease of application to the design of Weber and Crocker (1983). The RMC assigns exemplars sequentially to a cluster based on similarities in observed features using a Bayesian model to approximate the ideal partition:

$$p(k|f) = \frac{p(k)p(f|k)}{\sum_k p(k)p(f|k)} \tag{1}$$

where k is the cluster and f is the feature set of the exemplar under consideration. This posterior probability is calculated for all existing clusters as well as a new potential cluster, with the highest probability determining assignment. Following Anderson (1991), the prior probability was defined as:

$$p(k) = \begin{cases} \dfrac{cn_k}{(1-c)+cn} & \text{if } k \text{ is old} \\ \dfrac{(1-c)}{(1-c)+cn} & \text{if } k \text{ is new} \end{cases} \tag{2}$$

where n_k is the number of exemplars in cluster k, n is the total number of members assigned to the partition, and c is a coupling parameter describing the probability of two exemplars being grouped together independent of any observations.

The likelihood also followed the format of Anderson (1991):

$$p(f|k) = \prod_i p_i(j|k) \tag{3}$$

where the exemplar's features are divided into dimensions i holding values j. As stated above, the discrete form of this probability was used, in keeping with the exemplar structure used by Weber and Crocker (1983):

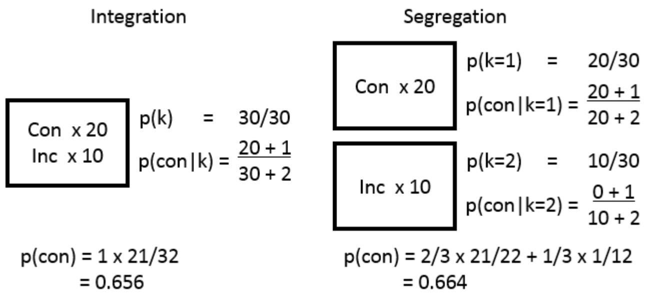
$$p_i(j|k) = \frac{n_j + \alpha_j}{n_k + \alpha_0} \tag{4}$$

where n_j is the number of exemplars in cluster k showing trait value j on dimension i, n_k is the number of cluster members showing a value on that dimension and α_j is a parameter reflecting the prior expectation of the occurrence of that value, while α_0 is the sum of these alpha values.

Once a partition has been generated, the model is then able to calculate a probability value measuring the likelihood of a new group member exhibiting a congruent trait value on any dimension. This is done by taking an average of the rate of congruent traits in each cluster weighted by the probability of that cluster:

$$p(con) = \sum_k p(k)p(con|k) \tag{5}$$

where $p(con|k)$ follows the format of Equation 4, focussing on congruent trait values. This explains how isolating incongruent data in a distinct subtype mitigates its impact: smaller clusters provide less evidence to outweigh prior expectations, here represented by the α parameter. As such, there is less confidence that future members of the incongruent cluster will demonstrate similar trait values, while the larger congruent cluster carries more certainty. To illustrate, consider a case in which 30 exemplars, 20 congruent and 10 incongruent, are either integrated or segregated. For the purposes of this illustration, $\alpha = 1$ for both congruent and incongruent traits, and $c = 1$, meaning no new cluster is considered:



As this shows, stereotype-congruency is estimated to be more probable in the segregated case because the α values are more impactful in the smaller cluster, offsetting the actual ratio of traits to a greater degree.

Stereotypes as Prior Clusters

As the category in question is a familiar social group with which participants are likely to have previous experience, the model included a cluster of exemplars added to the partition before exposure to the main exemplar set in order to simulate this prior knowledge. This also provided a more valid depiction of the origins of the group stereotype by making the members of this prior cluster stereotype-congruent, as well as allowing for potential interactions between prior knowledge and new information, as have been observed in other categorical modelling studies (Heit, Briggs, & Bott, 2004). Exemplars in the prior cluster therefore displayed stereotype-congruent values on all stereotypical dimensions, as well as group membership on a separate dimension, while the number of cluster members was added as an additional model parameter.

Alternative Models

While the above demonstration does show that the RMC is able to predict a subtyping effect, in order to determine whether this is truly the result of a rational process, the RMC must be compared with a more heuristic depiction of the subtyping mechanism. A second model was therefore developed in which the segregated incongruent data is ignored rather than simply mitigated, essentially redrawing category boundaries to exclude counter-stereotypical information. This was achieved by restricting the clusters considered when making probability estimates to only the cluster with the highest posterior probability. This restriction was based on the findings of Murphy and Ross (1994), which suggested that participants often only considered the most likely cluster when making probability estimates rather than all generated clusters. This creates an additional candidate model for assessment, labelled the Restricted Rational Model of Categorisation (RRMC).

In addition, as a counterpoint to this 'extreme subtyping' model, a third model was developed in which book-keeping was enforced; this was achieved by limiting the RMC to using only a single cluster by fixing the c parameter at 1, forcing all exemplars to be grouped together despite any differences in features. This 'Single Cluster Model' (SCM) therefore removes the possibility of segregating incongruent data, eliminating any influence of data format and focussing entirely on the ratio of traits in the partition.

These three models therefore present three different mechanisms of stereotype revision: while both the RMC and RRMC use a partition that flexibly adapts to observed data patterns, the RRMC subsequently simplifies this partition by focussing on only one cluster, heightening any effects this representation may have generated, while the RMC remains more moderate. Conversely, the SCM focuses on trait ratios rather than data pattern, thereby dismissing any effects that may be predicted by the other candidate models.

There is, however, a key distinction between these rational and heuristic models which can be used to determine their validity: in the RMC, the subtyping effect is dependent on the smaller size of the subtype cluster, meaning that increasing the size of the subtype by adding more incongruent members should reduce and ultimately eliminate this effect. In contrast, the RRMC will continue to ignore the subtype regardless of its size unless the subtype becomes so large that it is selected as the most likely cluster, at which point estimates will change drastically to reflect the subtype's much lower rate of congruency. This could essentially reverse the subtyping effect at higher volumes of incongruent evidence, focussing on counter-stereotypical rather than stereotypical clusters, and so bearing a closer resemblance to the conversion-effect described above. The SCM, meanwhile, is unable to exclude incongruent data at all, and therefore predicts no subtyping effect at any volume of incongruent information.

The accuracy of these models can therefore be contrasted according to the change in the subtyping effect with further exposure to stereotype-incongruent evidence: the RMC predicts a reduction in subtyping at higher volumes of counter-stereotypical data; the RRMC predicts a stable subtyping effect until a sudden reversal; and the SCM predicts no subtyping effect at any point. The following experiment therefore set out to compare these model predictions by extending the concentration design of Weber and Crocker (1983) across a higher total volume of evidence and taking measures of stereotypical beliefs throughout exposure. This also provided direct behavioural data for use in assessing the fit of the candidate models for a more complete test of these predictions.

Experimental Data

Method

Participants One-hundred-and-sixteen participants were selected from a University of Warwick undergraduate psychology class as part of a course requirement. The sample included 102 females and 14 males, while age ranged between 18 and 27 years, with a mean of 19.

Design and Materials The experiment followed the concentration design of Weber and Crocker (1983) with an additional within-subjects manipulation of data volume: measures of stereotypical beliefs were taken at fixed intervals during the observation of a set of exemplar descriptions where stereotype-incongruent information was either concentrated in a subset of exemplars or dispersed across all exemplars. Two exemplar sets were therefore created for use in the experiment, each containing 90 total exemplars displaying four trait dimensions: the first dimension described the occupational label, and so was identical for all exemplars, while the remaining three dimensions described personality traits with three possible values (stereotype-congruent, stereotype-incongruent or neutral). In both sets, two-thirds of the 270 total traits were incongruent, one-sixth were congruent and one-sixth were neutral; incongruent traits made up the majority in order to allow for a potential incongruent cluster to be larger than any other in the category. In the concentrated exemplar set, these incongruent traits were concentrated such that 60 exemplars each displayed incongruent traits on all three personality dimensions, with the congruent and neutral traits being distributed equally between the remaining exemplars. In the dispersed exemplar set, all traits were distributed as equally as possible.

As in Weber and Crocker (1983), exemplars were said to come from the category of lawyers; exemplars were therefore transformed into member summaries for use in the experiment by assigning each value on the three personality dimensions a unique trait label. Sixteen total labels were used: 5 congruent (Intelligent, Industrious, Neat, Out-going and Well-dressed), 5 incongruent (Incompetent, Lazy, Messy, Shy and Slovenly) and 6 neutral (Warm, Religious, Jovial, Obnoxious, Reserved and Meditative). These labels were taken from Weber and Crocker (1983), being based on pilot tests determining stereotypical and counter-stereotypical traits for the target category of lawyers. Three labels of each trait type were randomly selected at the start of each run of the experiment for use in exemplar summaries. Summaries were also assigned randomly selected names to assist in individuation.

Procedure Upon arriving at the lab, participants were first randomly assigned to one of the two concentration conditions, determining which set of exemplars would be viewed; this was balanced to provide equal numbers, meaning 58 participants were allocated to each condition. Participants were told the experiment tested how perceptions of a group changed with experience, involving both viewing summaries of group members and answering questions about the traits of the group in general.

The experiment began by asking participants to estimate the likelihood of certain traits appearing in the category of lawyers according to the number of members in a sample of 100 lawyers displaying that trait. Estimates were requested for all 16 possible personality traits, though only 9 were used in the subsequent member summaries. This first question block therefore provided a measure of baseline beliefs before any experimental exemplars were viewed.

After providing estimates for all traits, participants began a presentation block in which member summaries were shown on screen for the participants to examine. In order to maintain attention on this information, participants were asked to rate the pleasantness of each group member on a scale of 1-10, though this measure was not used during analysis.

At set intervals of presentation, the test block was repeated, and participants were again asked to estimate the likelihood of each of the 16 traits appearing in the category to measure any changes in expectation. This occurred after viewing 6, 18, 36, 60 and 90 total exemplars, with the ratio of traits within each interval being consistent with that of the complete exemplar set. At the start of each test block, participants were informed that though some of the questions had been asked before, they should answer based on how they felt at that point in time.

After viewing all 90 lawyer summaries and completing the final test block, the experiment ended, and participants were debriefed as to the aims and expectations of the study.

Results

Data Analysis

The results of the experiment were analysed using a mixed linear regression model including the factors of evidence volume, concentration condition and trait type. As the first test block was intended to provide a baseline, being unaffected by either volume or concentration, ratings from this round were not included in the regression model. This was confirmed using independent t-tests, finding no significant difference between conditions in either congruent ratings ($t(114) = .190$, $p = .850$) or incongruent ratings ($t(114) = .296$, $p = .768$) in the first test block.

The regression model showed significant effects for volume in both congruent ($\beta = -2.23$, $t(5086) = 6.41$, $p < .001$) and incongruent ratings ($\beta = 6.49$, $t(5086) = 15.6$, $p < .001$), with congruent ratings decreasing and incongruent ratings increasing over the task. Similarly, condition is shown to be a significant predictor for both congruent ($\beta = -6.36$, $t(114) = 2.50$, $p = .014$) and incongruent ratings ($\beta = 12.6$, $t(114) = 3.19$, $p = .001$), with congruent ratings being higher and incongruent ratings lower in the concentrated condition. Finally, the interaction between concentration and volume was found to significantly differ between congruent and incongruent ratings ($\beta = -2.21$, $t(5086) = 3.76$, $p < .001$), potentially indicating differences in the level of the subtyping effect over the task.

This was investigated further using two additional mixed linear regression models for each trait type, both including the factors of condition and evidence volume. Coefficient estimates from the congruent ratings model suggested evidence volume to be a significant predictor ($\beta = -3.11$, $t(1620) = 6.63$, $p < .001$), but concentration condition to be non-significant ($\beta = -3.63$, $t(114) = 1.24$, $p = .218$), with no significant interaction between these factors ($\beta = .58$, $t(1620) = .88$, $p = .379$). Conversely, the incongruent ratings model suggested a significant effect of volume ($\beta = 4.48$, $t(1620) = 7.79$, $p < .001$) and condition ($\beta = 7.40$, $t(114) = 2.40$, $p = .018$), with a near-significant interaction ($\beta = -1.52$, $t(1620) = 1.87$, $p = .062$). The findings of the general model are therefore most evident in the incongruent ratings when the two trait types are separated, while congruent ratings do not display such strong effects.

Model Comparison

Participant data was compared with model predictions made by the three candidate models to determine which provided the most accurate depiction of behaviour in the task. This used a grid search function across the three parameters, with the considered values being: for c, 0.01 to 0.99 in steps of 0.01; for α, 0.1, 0.2, 0.3, 0.5, 1, 2, 3, 5, 10, 15, 20, 25 and 30; and for membership frequency of the prior cluster, 0 to 50 in steps of 1. The models were run through the same exemplar sets given to participants at each combination of parameter

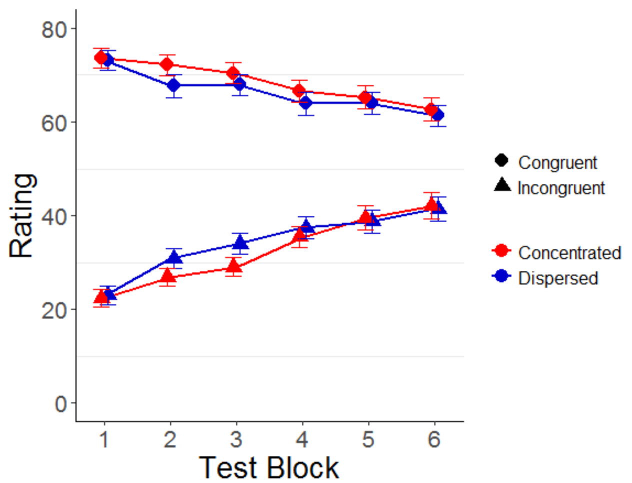


Figure 1: Trait ratings for both trait types in both concentration conditions across the 6 test blocks. Error bars show 95% CI.

values to generate estimates of the probability of both congruency and incongruency in new category members at each of the six exemplar intervals. These values were then used to calculate model likelihoods assuming identical parameter values for all participants in order to allow the model to fit both conditions simultaneously. Likelihoods were calculated in each of the six test blocks according to the product of the probability of all participant ratings for that trait type in that test block; these probabilities were defined according to a normal distribution using the model probability estimate as a mean and variance fit to maximise the final product. These values were then transformed into log likelihoods before being summed across test blocks and concentration conditions to create a single model log likelihood for all participants at that set of parameter values. Maximum log likelihoods from each model were then used to calculate BIC values for comparison. The RMC was found to have the lowest BIC score (11926, $\alpha = 0.5$, $c = 0.01$, prior membership = 21), indicating this model had a better fit to the experimental data than either the RRMC (11937, $\alpha = 10$, $c = 0.09$, prior membership = 50) or the SCM (11929, $\alpha = 10$, prior membership = 50).

Interestingly, when the predictions for this best fit for the RMC are examined, probability estimates for both measures are in fact identical between conditions; this is because all experimental exemplars were assigned to separate, single-member clusters despite any similarities in features, mitigating all exemplars equally in both sets. This suggests that the differences observed in participant ratings between concentration conditions were sufficiently small such that the data could be best fit by identical behaviour in both conditions. As previously described, this is in fact a tenet of the SCM, which ignores the concentration of data via full integration of all exemplars; however, the SCM shows a steeper curve in both measures compared to the RMC, therefore predicting greater stereotype revision. As such, the scattering behaviour of the RMC better corresponds with the greater degree of maintenance observed in the data.

It is also notable that the best fit of the RRMC matches that of the SCM, as the maximum likelihood of the RRMC was found when all exemplars were grouped in a single cluster.

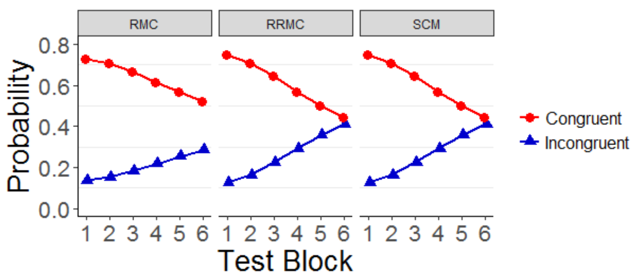


Figure 2: Trait probability estimates from the best fits of the three candidate models. Due to equality in estimates between conditions for all models, only one line is used for each measure.

As such, the RRMC also predicts greater revision than was observed in the experiment; however, because the SCM does not use a coupling parameter, the SCM holds a lower BIC value than the RRMC despite equal log likelihoods. It should be noted however that this comparison reveals only the best fit of the three candidate models rather than an absolute description of behaviour in the task; more complex models may therefore be needed to reflect the subtle differences observed in the participant data.

Discussion

The results of the experiment provide three key findings: firstly, ratings of trait likelihood for both congruent and incongruent traits became less stereotypical over the course of the experiment, indicating that higher volumes of incongruent evidence were effective in evoking greater revision of stereotypical beliefs. Secondly, ratings were more stereotypical in the concentrated condition compared to the dispersed condition, as would be expected by subtyping. Thirdly, this concentration effect differed somewhat in size across the task, showing smaller differences between groups at higher volumes of evidence. When the trait types are separated, these findings are seen to be stronger in the incongruent ratings, while congruent ratings did not demonstrate the condition or interaction effects.

In general, these results appear to partially correspond with previous depictions of the subtyping effect: beliefs are more stereotypical where incongruent information is more easily segregated from existing category structures, whereas data patterns aiding integration demonstrate greater stereotype revision. This also matches with the categorical explanations for subtyping offered by both the RMC and RRMC, as both suggest that category partitions which place incongruent data in a separate cluster diminish the impact of this data on subsequent probability estimates, thereby leading to more stereotypical expectations.

There is, however, an additional aspect to the subtyping effect observed in this task which distinguishes between these models: the interaction between volume and concentration, while not quite reaching a significant effect in the separated regression models, does suggest that the subtyping effect did not remain entirely consistent across the task, but in fact dropped off in later test blocks, with incongruent trait ratings in particular appearing to converge between the two conditions. This finding corresponds with the predictions of the RMC made in the introduction to this study: because the subtyping effect in the RMC is the result of greater uncertainty in the data pattern of the subtype cluster due to its smaller size, increasing the size of this cluster attenuates the subtyping effect by providing more confidence in this pattern. This is in contrast to the RRMC's hypothesised cross-over from subtyping to conversion at higher volumes of incongruent evidence where the subtype becomes the most likely cluster due to its size, an effect which was not observed in the data.

The RMC therefore appears to provide the most accurate theoretical account of the observed results, suggesting participants were most likely using a rational categorisation process to guide their judgements in this task. This suggestion is further supported by the RMC having the best fit to the behavioural data in the above model comparison, though it is notable that this best fit did not accurately capture the observed differences between concentration conditions. Even so, the RMC does still provide a better fit to the experimental results than either the RRMC or SCM, potentially suggesting the present findings are more likely to be due to standard rational processes than these more extreme depictions.

Implications

The present study therefore provides evidence from both behavioural data and model fits that the maintenance of stereotypical beliefs generated by subtyping does appear to be the result of a rational incorporation of all available data rather than a heuristic strategy of stereotype preservation: isolating incongruent data in a distinct subtype does not completely exclude this information from consideration during category judgements, but instead mediates its impact according to the size of the subtype.

As such, the subtyping effect could be considered to be a normal aspect of standard categorisation processes operating on social groups, occurring where a particular data pattern inadvertently diminishes the impact of counter-stereotypical data. If so, stereotype change could be encouraged by using similarly rational techniques to circumvent subtyping, primarily by aiding the integration of incongruent data into pre-existing clusters. More broadly, this finding provides a basis for a rational system underlying stereotyping, allowing for the generation of further predictions regarding stereotype use based on the principles of such rational models to be tested in future studies; such tests would be valuable in further developing the current model to provide a more complete depiction of rational stereotype use.

The current data also demonstrates that stereotype change can be drawn from even slight encounters with incongruent evidence in sufficient volume, with the effects observed in the experiment being based solely on the observation of member summaries rather than any significant interaction with actual counter-stereotypical group members. This is in contrast to past theories such as the Contact Hypothesis which often require intensive, long-term interaction with out-group members to generate a reduction in stereotypical beliefs (Allport, 1954). This is not to say that prior expectations have been completely overcome: revision still does not reach the level suggested by conversion (or indeed the actual ratio of evidence in the experimental data sets); even so, this does still provide limited evidence that stereotype maintenance can be counteracted through increased exposure to incongruent data.

The current design may therefore present a more economic path to combating prejudice, requiring less time and effort than some existing methodologies. What is more, the effects observed in this study could in fact be greater at more significant forms of encounter, potentially counteracting subtyping at even lower volumes of incongruent data. It is not clear how the significance of an encounter should be represented within the current version of the model, but one basic option would be to represent a significant encounter as multiple observations in the partition, essentially viewing that individual as providing more data than a single exemplar. This suggestion should, however, be pilot tested to determine the validity of this representation before being incorporated into the model.

Conclusion

The present study provides the starting point for a rational approach to stereotype use, providing both theoretical and empirical evidence that a rational model of stereotype change, while not universally accurate, does provide a reasonable account of behaviour both in this experiment and previous studies into stereotype maintenance. We therefore hope that this study can act as a foundation for continued work in this field, allowing subsequent research to further refine the presented models to provide a more accurate depiction of behaviour. This will serve to provide greater clarity regarding the operations underlying stereotype maintenance, and so aid in finding more potential methods for encouraging stereotype change.

References

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, Mass: Addison-Wesley.

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

Bott, L., & Murphy, G. L. (2007). Subtyping as a knowledge preservation strategy in category learning. *Memory & Cognition*, *35*, 432–443.

Heit, E., Briggs, J., & Bott, L. (2004). Modeling the effects of prior knowledge on learning incongruent features of category members. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 1065–1081.

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

Johnston, L., & Hewstone, M. (1992). Cognitive models of stereotype change: 3. subtpying and the perceived typicality of disconfirming group members. *Journal of Experimental Social Psychology*, *28*, 360–386.

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

Weber, R., & Crocker, J. (1983). Cognitive processes in the revsion of stereotypic beliefs. *Journal of Personality and Social Psychology*, *45*, 961–977.