

Illusory Correlation, Group Size and Memory

Dirk Van Rooy, Australian National University, Australia

Tim Vanhoomissen, Lessius Hogeschool, Belgium

Frank Van Overwalle, Vrije Universiteit Brussel, Belgium

Address for correspondence: Dirk Van Rooy, School of Psychology, Australian National University,
ACT 0200, Australia; or by e-mail: dirk.vanrooy@anu.edu.au.

Running Head: Illusory Correlation and Memory

[PUBIC]

Number of words: 7613

21 December 2012

Abstract

Two studies were conducted to test the predictions of a multi-component model of distinctiveness-based illusory correlation (IC) regarding the use of episodic and evaluative information in the production of the phenomenon. Extending on the standard paradigm, participants were presented with 4 groups decreasing in size, but all exhibiting the same ratio of positive to negative behaviours. Study 1 (N = 75) specifically tested the role of group size and distinctiveness, by including a zero-frequency cell in the design. Consistent with predictions drawn from the proposed model, with decreasing group size, the magnitude of the IC effect showed a linear increase in judgments thought to be based on evaluative information. In study 2 (N = 43), a number of changes were introduced to a group assignment task (double presentation, inclusion of decoys) that allowed a more rigorous test of the predicted item-specific memory effects. In addition, a new multilevel, mixed logistic regression approach to signal-detection type analysis was used, providing a more flexible and reliable analysis than previously. Again, with decreasing group size, IC effects showed the predicted monotonic increase on the measures (group assignment frequencies, likability ratings) thought to be dependent on evaluative information. At the same time, measures thought to be based on episodic information (free recall and group assignment accuracy) partly revealed the predicted enhanced episodic memory for smaller groups and negative items, while also supporting a distinctiveness-based approach. Additional analysis revealed that the pattern of results for judgments thought to be based on evaluative information was independent of interpersonal variation in behavioral memory, as predicted by the multi-component model, and in contrast to predictions of the competing models. The results are discussed in terms of the implications of the findings for the proposed mechanisms of illusory correlation.

Humans sometimes show an unsettling insensitivity for actual relationships between events and often seem to perceive a covariation when there is none, a bias that has been labeled *illusory correlation (IC)*. This bias has important practical consequences as it has been suggested that it contributes to stereotyping (Hamilton, 1981), faulty clinical judgments (Dawes, 1989; Chapman & Chapman, 1967), depression (Seligman, 1975) and minority discrimination (Hamilton & Gifford, 1976). The paradigmatic demonstration of this IC bias in social psychology comes from a study on the formation of group stereotypes by Hamilton and Gifford (1976). In their study, participants read behavioral statements about members of a majority group, labeled A, and a minority group, labeled B. Both groups revealed the same ratio of desirable to undesirable behaviors (9:4), but twice as many statements referred to members of group A than to members of group B. As such, there was no objective correlation between group membership and desirability of behavior. Nevertheless, after reading all statements, participants showed greater liking for the majority group A than for the minority group B. In addition, perceivers overestimated the frequency with which members of the minority group B had engaged in the less frequent undesirable behaviors. This result demonstrated that a simple asymmetry in information sampling can lead to differential perception of social groups.

Distinctiveness Account

The most dominant theoretical explanation for IC in the last decades has been the distinctiveness-based explanation (DBE; Hamilton & Gifford, 1976). Its basic premise is that infrequent or otherwise salient items are encoded more extensively at exposure and are therefore more accessible later, when judgments are made. According to this reasoning, in a typical IC experiment, the small number of minority and undesirable behaviors makes them more salient or distinctive, leading to an overestimation of the frequency of these behaviors and less liking of the minority group B.

A number of studies seem to support this premise. Hamilton, Dugan and Trolhier (1985) showed that IC only occurs if information is presented in a typical sequential fashion (statement per statement) and not when information is presented in a summary table that tends to attenuate the saliency of undesirable minority (or B-) behaviors. Further evidence for the enhanced encoding of

distinctive items comes from a study by Stroessner, Hamilton and Mackie (1992), who found that participants spend more time reading distinctive B- items than the other items. Several studies also found that B- items are better retrieved in free recall tasks (Hamilton et al., 1985; Stroessner et al., 1992; McConnell, Sherman & Hamilton, 1994) and assigned faster to their group than the other behaviors (Johnson and Mullen, 1994; McConnell et al., 1994). Hamilton et al. (1985) found a high correlation between recall of group B behaviors and evaluative judgments of group B, indicating that the high accessibility of group B behaviors influenced evaluative judgments about this group. Most of the empirical support for the basic assumption of the DBE, namely that IC results from distinctive items being more extensively encoded and more accessible in memory, has been obtained in experiments using only one large and one small group. In a notable exception, Sherman, Hamilton and Roskos-Ewoldsen (1989) found that including a third group C that was smaller, identical or larger in size than group B diminished, but did not eliminate the IC effect involving group B. Importantly, this was the case both when B- behaviors were the most distinctive (when group C was larger or identical in size) as well as when they were not (when group C was smaller) - a finding that clearly challenges the original DBE. The need for a more conclusive definition of "distinctiveness" was further illustrated by a recent study by Risen, Gilovich & Dunning (2007): In a number of studies, they revealed the emergence of one-shot illusory correlations, by showing that a single instance of unusual behavior by a member of a rare group is sufficient to create an association between group and behavior. Using real groups, their results showed that single unusual behaviors committed by members of rare groups were processed differently than other types of behaviors, and were more memorable. In sum, these studies suggest that a more thorough exploration of how distinctiveness drives IC is in order. Our studies are, in part, designed to do this.

Alternative Accounts

A number of authors have suggested alternative accounts of IC that differ drastically from the original DBE. Fiedler's Information Loss Account (Fiedler, 1996; Fiedler, Kimmelmeier, & Freytag, 1999) states that illusory correlation results from better extraction of information about the majority than

minority group, leading to greater regression to the mean in perceptions of the positive-to-negative ratio of the minority group. More specifically, it is argued that group judgments are based on traces that are retrieved from stored exemplars in memory and then aggregated, according to a simple weighted linear summation. This aggregation process cancels out unsystematic information or encoding errors and so reinforces systematic tendencies. It is sensitive to group size: as the amount of observations on which decisions are based increases, less error variance is left in the aggregate, rendering perceptions of the group more accurate and leading to less biased judgments (see Fiedler, 1996, p. 200-201, section on "*Illusory correlation simulation*"). Within the typical IC design, this process results in an evaluative bias that favors the majority group, as its larger group size allows better reproduction of its aggregate, evaluative characteristics (i.e. more desirable than undesirable information). Because of its smaller size, the true ratio of positive to negative behaviors is learned less well for the minority group, leading to a more extensive regression to the mean in judgments about the minority group. A number of studies (Fiedler, 1991; 1996) have supported this account, basically showing the occurrence of IC in the *absence* of distinctive behavior and *without better memory* for infrequent events.

McGarty, Haslam, Turner and Oakes (1993) have suggested that IC results from a differentiation process, in which perceivers accentuate differences between the groups based on the valence of the behaviors (see also Berndsen & Spears, 1997). Given that there is more information on the positive behaviors of group A, there is more evidence for the hypothesis that Group A is better than Group B, than for the opposite hypothesis. This forms the basis for further accentuating the apparent evaluative differences between the groups, leading to an IC effect (Berndsen & Spears, 1997; Haslam, McGarty & Brown, 1996; McGarty et al., 1993). According to this perspective, strictly speaking, the distinction that achieves the greatest differentiation involves assigning all positive behaviors to group A and all negative behaviors to group B. A number of studies have provided support for this account. It has been found that the IC disappears when two well-known groups differ on a non-evaluative dimension that does not provide a meaningful (i.e., clear and separable) distinction between behaviors and groups, such as right- versus left-handedness (Haslam et al., 1996) or group members' gender

(Klauer & Meiser, 2000). The constructive nature of the categorization process has been illustrated by data showing that participants actively interpret the behavioral statements over the course of the stimulus presentation in light of the developing group differentiation and that the IC develops and strengthens towards the end of the experiment (Berndsen, Spears, McGarty & van der Pligt, 1998).

Recently, Sherman and colleagues (2009) proposed a theoretical framework based on Attention Theory (Kruschke, 2001, 2003) that integrates a number of the previous accounts. Their model extends the traditional distinctiveness explanation, by positing that differentiation occurs via a focus on *contextually* distinct information. Because there is more information about the majority group, features of the majority group are learned before features of the minority group. The features of the minority group that subsequently stand out, are those that most distinguish it from what has been established about the majority group, serving to further differentiate the groups. Consistent with this, they found that participants paid more attention to common (positive) than rare (negative) behaviors when reading about majority group members, but rare traits received more attention when reading about minority group members (see Sherman et al., 2009, Experiment 5). So as with the distinctiveness account, special attention is paid to infrequent behaviors performed by members of the minority group. However, unlike that account, and more in line with the differentiation account, this model proposes that the basis for that attention is contextual distinctiveness in relation to the majority group, rather than absolute numerical distinctiveness. As such, the Attention Theory model, or AT model, is a combination of a distinctiveness and differentiation approach.

Multiple component model

Each of the above explanations has received considerable empirical support, where support for any particular account over others is typically found with some manipulations and measures, but not with others. The models of Fiedler (1991) and McGarty *et al.* (1993) share the common feature that they do not give any special status to small group rare behavior events, and thus have trouble accounting for evidence indicating enhanced memory for such events. The distinctiveness explanation, on the

other hand, has trouble accounting for the occurrence of ICs in the absence of superior memory for these events. As such, the picture that emerges is that of illusory correlation as a complex, multiply determined phenomenon. A number of authors have therefore suggested that it might be useful to distinguish between different components underlying IC (Johnson & Mullen, 1994; Johnson, Mullen, Carlson & Southwick, 2001). This idea was further extended and implemented in a connectionist model of group processes by Van Rooy et al. (2003), who applied it to a number of group biases, including IC. According to this model, to which we will refer as the Multiple Component Model (MCM), perceivers create a mental representation containing connections of a social group with a global evaluative impression, and with episodic behavioral information. More particularly, perceivers are thought to incrementally develop prototypes of both the majority and minority group that contain mainly evaluative (desirable or undesirable) information, but also store item-specific behavioral information (an episodic trace). With every statement presented, that prototype is updated using a connectionist learning algorithm (i.e., the delta algorithm by McClelland & Rumelhart, 1988). For instance, if the information describes a member of group A performing a desirable behavior ("John, member of group A, helps an old lady across the street"), the connection between the group A and a desirable evaluation becomes stronger. At the same time, the specific behavior is encoded in an episodic trace that is connected to group A as well. Thus, each piece of behavioral information is encoded at two levels: Its evaluative meaning ("the behavior is good") and its unique episodic meaning ("helps an old lady across the street").

Due to their separate representation, the MCM makes different predictions for judgments based on evaluative versus episodic information. On judgments that do not require the retrieval of item-specific episodic information (likeability ratings of the group, frequency estimation and assignment of behaviors to groups), the model predicts an evaluative bias in favor of the majority group. The reason is that more information is added so that the evaluative connections with the group grow stronger. Crucially, increasing connections on the group level impairs memory for discrete episodic traces of information. This prediction is based on a principle of competition between the weights of the connections, so that stronger (evaluative) connections tend to block the development of weaker (episodic) connections (see Van Rooy et al., 2003, p. 540, section on "Competition

property and discounting”). This principle is akin to the ideas of discounting in causal explanation (Kelley, 1971) and associative learning (Rescorla & Wagner, 1972), which say that stronger explanations tend to overshadow weaker ones. Because of its greater group size, this impairment of episodic connections is more detrimental for the majority group A as compared to the minority group B, leading to better retrieval of minority behaviors. Also, because there are typically more desirable behaviors than undesirable in an IC experiment, stronger memory impairment occurs for desirable behaviors, leading to better retrieval of undesirable behaviors (i.e., B- behaviors). These memory effects are predicted to be revealed in accuracy measures of episodic traces, such as accuracy of group assignment and free recall. Van Rooy, Van Overwalle et al. (2003) have illustrated how their model can account for a number of critical IC findings, and its basic tenet that illusory correlation effects reflect incomplete learning rather than a bias due to information loss was recently confirmed by Murphy, Schmeer, Vallée-Tourangeau, Mondragón, & Hilton (2011).

The Current Studies

In two studies, we set out to test the predictions of the MCM model of how different types of information in a classic IC design contribute to produce the typical IC effects on different dependent measures. We chose to focus on the MCM, not because it is superior to other accounts, but rather because it generates predictions about how the application of episodic and evaluative information evolves in function of sample size. No study to date has tested these predictions directly, and a number of modifications were applied to the standard IC design and its analysis to make this possible.

The first difference with previous studies is the use of four instead of two groups, decreasing in size while maintaining the usual 2-to-1 ratio of desirable to undesirable behaviors. Although previous studies have extended the IC design in comparable ways (see Fiedler, 1991; Sherman et al., 1989), none of these studies used appropriate memory measures (see below) to analyse episodic item memory. The MCM predicts that, as group size decreases, a monotone increase in ICs should

be apparent in evaluative measures for which participants rely on a general group impression, rather than specific items (likability ratings, frequency estimation, group assignment task). In contrast, in terms of episodic information, the MCM predicts that item memory should improve as events become less frequent. In other words, it predicts that real item-memory should improve as group size decreases, and a memory advantage for (less frequent) negative events should become apparent.

To test the MCM predictions regarding episodic memory, we will apply a novel multilevel approach to signal detection analysis. We chose signal detection analysis, as a number of authors have illustrated that it performs better than threshold models (see Klauer & Meiser, 2000), particularly in unbalanced designs (Slotnick & Dodson, 2005; Slotnick, Dodson, Klein & Shimamura, 2000; see Method section), although Klauer and Kellen (2010) have contested Slotnick et al.'s arguments. This analysis enables us to reliably test the MCM prediction that enhanced behavioral memory for smaller groups is revealed in judgments driven by item-specific, episodic information, and this quite independently from the bias apparent in judgments based on evaluative information.

Experiment 1: Group size and evaluative information

In this experiment, we set out to test our main group size manipulation and further explore the role of distinctiveness. As mentioned, the MCM predicts a monotonic increase in illusory correlations in function of decreasing sample size, which should be apparent in the likeability ratings, frequency estimation tasks and the group assignment data. To investigate the role of distinctiveness in the current design, we included a zero-frequency cell by removing all undesirable behavioral statements from the minority group D. A similar manipulation has been applied by other studies (Fiedler, 1991; Shavitt, Sanbonmatsu, Smittipatana & Posavac, 1999), but not in combination with the current group size manipulation and memory analysis, which are critical in testing the MCM predictions. The MCM predicts a monotone decrease in evaluative ratings in function of a smaller group size, regardless of the empty cell. This is because at the level of general, evaluative information, decreasing group size will result in impaired encoding of the groups' evaluative characteristics, leading to stronger IC

effects. At the same time, the MCM predicts stronger memory for discrete episodic traces in function of decreasing group size.

Method

Participants

Participants were 75 female and male freshmen from the Dutch speaking Vrije Universiteit Brussel, who participated for a partial requirement of an introductory psychology course. They were tested individually in cubicles.

Materials

In a preliminary study, a list of 100 Dutch statements describing everyday behaviors ("John helps an old lady across the street", "Albert is very modest") was presented to a group of 70 students that rated each statement on a five-point scale in terms of desirability. Hamilton and Sherman (1989) showed that the typical IC effects are most likely to occur when participants are presented with moderately desirable statements. Therefore, only statements with moderate mean values of desirability and undesirability ($0.5 < M < 2.5$ for undesirable and $3.5 < M < 4.5$ for desirable behaviors) were selected for use in the experiment. This resulted in an experimental set of 45 statements, consisting of 30 moderately desirable behaviors and 15 moderately undesirable behaviors.

For each participant, these statements were randomly assigned to four groups in accordance to the frequencies in Table 1. The ratio of desirable to undesirable behavior was 2 to 1 in every group, so there was no objective relationship between group membership and evaluative character of the behaviors. To avoid any gender bias, names ("John") were replaced by initials ("J.") in the actual experiments. This means that each statement presented to a participant consisted of an initial, a group designation (A, B, C or D), and a behavior. For example: "J., member of group A, helped an old lady across the street".

Procedure

Participants sat at individual computers and instructions appeared on the PC screen. They were informed that the experiment concerned “the way people process and retain information” and that they would receive information concerning four groups, A, B, C and D. They were told that these represented groups in the real world and that group A was bigger than group B, group B bigger than group C and so on. Finally, they were instructed to read each statement carefully. Each statement remained on the screen until the participant pushed the space bar. After reading all statements, participants completed a filler task that lasted about 3 minutes. In this filler task, a number appeared on the screen and participants had to press the corresponding key on the number pad as soon as possible.

After the filler task, participants completed a number of tasks in the following order (Because no effect of group labels or order of tasks were found in pilot studies, these manipulations were left out of the design of this experiment).

Free recall. Participants were provided with a blank piece of paper and were instructed by the computer to write down as many of the statements as they could recall, including the group membership of the actor that engaged in the behavior. If they could not remember group membership, they were encouraged to guess. After 20 minutes, the experimenter collected the papers. In all the experiments, two judges, blind to experimental manipulations, scored free recall data by using a gist criterion. In case of dispute, the decision of a third judge prevailed. Note that group labels were ignored in scoring accuracy of the recalled statements, because this makes it a more adequate measure of pure episodic (behavioral) memory.

Likeability rating. Participants were presented with 5 positive and 5 negative adjectives. For each group they were asked to indicate to what degree each of these adjectives was appropriate to describe members of that group on a 10-point scale ranging from *not appropriate at all* (1) to *very appropriate* (10). These 10 ratings were averaged to represent the perceived likeability of the group.

Group assignment. The 45 statements, without group membership, were presented again to the participants in a random order. Participants indicated as fast as possible which group member they thought had engaged in the behavior by pressing one of four keys on the keyboard.

Frequency estimation. Participants were given the total amount of statements for each group and were then asked to indicate how many of these behaviors were undesirable. After completing these measures, participants were debriefed, thanked for their cooperation and urged not to share any information concerning the experiments with other students.

Data Analysis

As noted earlier, we apply a novel multilevel approach to signal detection analysis. The norm within most memory recognition research is to calculate hit and false alarm rates for each participant and within-participant condition, and then to compare these aggregate measures. However, given the small and unequal cell sizes in our design, such an approach would produce unreliable estimates. Instead, we introduce a mixed logit model analysis (i.e., multilevel logistic regression) that provides an alternative way of performing signal detection analysis by moving away from analyzing aggregate level data (see also DeCarlo, 1998). The use of a regression model also provides an added degree of flexibility in analyzing the impact of continuous variables, such as reading time. But most importantly, mixed models have been shown to avoid spurious effects better, and have more power than traditional ANOVA analysis (Jaeger, 2008). See Appendix for more details.

Results

Group Impression

The results are listed in the top panel of Table 2. Overall, the results show the predicted pattern for group impression (i.e., likeability ratings, frequency estimates and group assignment). Note that the group assignment measure indicates how desirable versus undesirable behaviors are attributed to a group, and thus reflects general group impressions. The results indicate that participants estimated the amount of undesirable behaviors in groups A and D to be equal. This clearly contradicts the DBE, which predicted a higher estimate for group C. Interestingly, this also indicates that participants

overestimated the frequency of undesirable behaviors in group D, which were actually zero. The data were statistically analyzed with an analysis of variance (ANOVA) with Group Size (A to D) and, when appropriate, Desirability (desirable versus undesirable behavior) as within-participants factors. Apart from an omnibus ANOVA, we also conducted contrast analyses to see whether any effect of Group Size was linear, as we would predict. The critical effects of Group Size and linear trends are reported in the table.

Likeability ratings. As can be seen in Table 2, the likeability ratings reveal the predicted decrease of likeability for smaller groups, and this decrease was linear.

Frequency estimates. As predicted, there was a significant increase of estimated undesirable behaviors, although this increase was not linear. Post-hoc contrast analyses show that participants estimated the proportion of undesirable behaviors to be smaller for groups A and D than for groups B and C, $F(1, 74) = 13.27, p < .001$. As noted earlier, there was no difference in estimates between groups A and D, $p > .9$, despite the actual absence of any undesirable behavior in group D.

Group assignment. As can be seen in Table 2, an omnibus ANOVA revealed a main effect of group size. As the actual proportions (between parentheses in the table) are already showing a linear decreasing pattern, a test for linear trend is theoretically not very meaningful for this measure. Of more interest are the deviations from that pattern, specifically, whether desirable and undesirable behaviors differ within groups. Consistent with predictions, contrast analyses show that participants tended to assign more desirable than undesirable behaviors to group A, $F(1, 74) = 13.54, p < .001$. In contrast, as predicted, participants tended to assign more undesirable than desirable behaviors to group B, $F(1, 74) = 3.38, p < .07$, and group C, $F(1, 74) = 8.31, p < .001$. There was no significant difference in assignments for group D, $p > .70$, which again indicates that participants wrongly assigned undesirable behaviors to group D (although there were none).

Item-Specific Memory

To analyze item-specific memory, we subjected not only free recall data, but also the group assignment data to an accuracy analysis, using multi-level analysis. Unlike the standard group

assignment analysis, which reflects general group impressions, the accuracy analysis should reflect true memory.

Free recall. The results are shown in the bottom panel of Table 2. Overall, the memory measures were generally quite weak for free recall. The structural zero D-cell makes an omnibus ANOVA not possible for these measures. Hence, separate ANOVA's were conducted for desirable and undesirable behaviors. A marginally significant effect of group size was found for desirable behaviors. This was mainly due to the higher recall of D+ behaviors as compared to recall of the other desirable behaviors, $F(1, 70) = 3.40, p = .07$. There was no significant effect of group size in recall for undesirable behaviors. As predicted, undesirable behaviors were generally better recalled than desirable behaviors, $F(1, 70) = 26.22, p < .001$.

Group assignment accuracy. Multilevel logistic regression analysis revealed that assignment accuracy shows a decreasing linear trend as groups become smaller, $\beta = -.24, p < .01$. Moderator analysis revealed a significant interaction between valence and linear trend, $X^2(1) = 13.57, p < .001$, indicating that accuracy for positive and negative items developed differently in function of group size. Further tests revealed that accuracy decreased for positive items, $\beta = -0.32, p < .001$, but not for negative items ($p > .3$). There was no significant change in accuracy for C- items ($p > .9$). We also failed to find a significant relationship between reading time and accuracy.

Discussion

In general, the results of Experiment 1 confirm the hypothesized impact of group size on IC effects. Making group size smaller across four groups, with the same 2:1 ratio of positive to negative behaviors, led to a linear decrease in positive, evaluative judgments across the groups. This linear trend is fully in line with the MCM, and clearly not compatible with any notion of special distinctiveness for a particular group. This interpretation is further strengthened by the occurrence of significant IC effects for the smallest group D in the *absence* of D- behaviors. Importantly, although a key prediction of both the MCM and ILA, this is in fact the first study that clearly reveals this predicted linear trend.

With respect to item-specific memory, the analysis of group assignment accuracy revealed a significant decrease with decreasing group size for desirable items, while there was no change for undesirable items. This is clearly inconsistent with the MCM, which predicted an increase in assignment accuracy, and is more in line with an information loss account (Fiedler, 1991). However, the results obtained for free recall were less straightforward. We found a marginally significant increase in the recall of desirable behaviors for the smallest D group compared to the other three groups, and better overall recall for undesirable as compared to desirable behaviors. Both of these results provide some support for the MCM and a distinctiveness account. However, the generally low level of memory performance and absence of distinctive negative items for group D, suggest that these results should be interpreted with some caution.

Experiment 2: Increasing Episodic Memory

Some aspects of item-specific memory performance in Experiment 1 were inconclusive. Therefore, in Experiment 2, we made two adjustments to achieve more reliable memory performance. First, we modified the group assignment task so as to isolate and eliminate evaluative guessing biases as much as possible. Given that group assignments are susceptible to IC, it is highly probable that participants use the desirability of behavior to guess that the statement came from group A. Given the high prevalence of positive A behaviors, this strategy may be quite successful. To identify these guessing biases in the group assignment task, decoy items were added that were identical to the original statements, but reversed in valence. For instance, instead of stating that an actor “helped the old lady across the street”, in these decoys it is said that the actor “did not help the old lady across the street” (for a somewhat different type of “distractors”, see Klauer & Meiser, 2000; Fiedler, 1993). Assigning these decoy items to a group indicates that a guessing strategy was used based on the behavior’s desirability, rather than true memory. In addition to an accuracy analysis, we can also analyze whether participants’ ability to distinguish between an old item (i.e., an item that was presented previously) and a decoy was affected by group size – in other words, whether participants were more accurate in recognizing old items from smaller groups. If memory is indeed enhanced for

smaller groups, as predicted by the MCM, such a pattern should be revealed in the ability to distinguish between old items and decoys.

Second, all statements containing behavioral information about the groups were presented twice, which we hoped would enhance free recall and lead to stronger memory in both the group assignment and free recall task. Presenting items twice is, as far as we know, a new procedure in the illusory correlation literature (but not in person memory, see Hamilton et al., 1981). As mentioned earlier, the MCM predicts better memory at the level of item-memory for smaller groups and for undesirable behaviors.

Method

Participants

Participants were 43 female and male freshmen from the Dutch speaking Vrije Universiteit Brussel, who participated for a partial requirement of an introductory psychology course. They were tested individually in cubicles.

Materials and Procedures

The same stimulus materials and procedure were used as in Experiment 1 with the following exceptions. First, the complete set of statements was presented twice, separated by a pause screen saying that "all statements will be presented for a second time". Second, in the group assignment task, not only the 45 original statements, but also 45 decoy statements were randomly presented. These items were identical to the original items, but reversed in valence. Participants were asked to press one of four buttons to indicate which group they thought the behavior came from, or a fifth button if they thought it was an item they had not seen before.

Results

Group Impression

The results and the F-values of Group Size and the linear trend analyses are depicted in the top panel of Table 3. Overall, the results confirm the decreasing pattern in group impression predicted by the alternative models:

Likeability ratings. There was a significant decrease in likeability for the smaller groups and this trend was linear.

Frequency estimates. There was a significant increase in the estimated proportion of undesirable behaviors for smaller groups, and this trend was also linear.

Group assignment task. An omnibus ANOVA revealed a main effect of Group Size (see Table 3) and Desirability, $F(1, 42) = 258.71, p < .001$, as well as a significant interaction, $F(3, 126) = 39.15, p < .001$. As predicted by the alternative models, subsequent contrast analyses showed that more desirable than undesirable behaviors were assigned to group A, $F(1, 42) = 110.86, p < .001$, as well as for group B although to a smaller degree, $F(1, 42) = 4.15, p < .05$. There was no significant difference in assigning desirable and undesirable behaviors to group C and group D.

Item-Specific Memory

The results are shown in the bottom panel of Table 3. As we had hoped for, the memory measures were much stronger, allowing us to perform a more reliable analysis.

Free Recall. An omnibus ANOVA revealed that free recall was better for smaller groups, and this trend was linear. No other effects were significant.

Group assignment accuracy. Multilevel logistic regression analysis revealed that assignment accuracy increased as groups became smaller, $\beta = .08, p < .05$. Adding a term testing for a quadratic trend did not improve the fit of the model ($p > .07$). Further tests showed that, compared to negative items, accuracy was reduced for positive items, $\beta_2 = -0.72, p < .001$. Moderation analysis did not reveal any significant interaction effects. As in experiment 1, we failed to find a significant relationship between reading time and accuracy.

Decoy analysis. The inclusion of decoy items allowed us to apply a more elaborate signal-detection type of analysis, in which we used participants' ability to distinguish between originals and decoys as a measure of memory sensitivity (see Appendix for more details). These analyses revealed that, as one would expect, participants respond "old" more to items they have seen before, rather than to decoys, $\beta = 2.21$, $p < .001$. There was also a main effect of valence. Compared to negative items, participants were in general more likely to categorize positive items as "old" rather than decoys, $\beta = .51$, $p < .001$. Moderator analysis revealed a significant interaction between valence and whether the item was old or a decoy, $X^2(1) = 5.23$, $p < .02$, showing that participants were more accurate for negative items, $\beta = 2.29$, than positive items, $\beta = 1.87$. Further moderator tests revealed a marginal linear ($p \approx .5$) and slightly stronger increasing quadratic trend ($p < .05$) in memory performance. Follow-up analysis revealed that this was mainly due to an increase in memory sensitivity for items from group D. This was apparent in the interaction effect between whether the item was old or a decoy, and whether or not it came from group D, $X^2(1)=11.26$, $p < .001$. This effect was further moderated by valence, $X^2(1) = 15.33$, $p < .001$, revealing that participants were more accurate for negative D items, $\beta = 2.30$, than for positive D items, $\beta = 0.22$.

Discussion

In Study 2, the predicted effect of increasingly negative group impressions given smaller groups was again observed in the likeability ratings, frequency estimates and group assignments. In addition, memory for behavioral information improved compared to Study 1: Both free recall data and group assignment accuracy revealed that episodic memory improved for smaller groups, which is in line with MCM predictions. The decoy analysis provided some further support for this, by revealing a marginal, increasing linear trend in memory performance. Follow-up analysis revealed that this was mainly due to improved performance for negative D items. However, contrary to a distinctiveness-based explanation, this superior memory for D- items did not affect the results on the other dependent measures (likeability ratings, frequency estimates and group assignments), which showed the same monotone increase in negativity in function of decreasing group size as in study 1. This

strongly suggest that participants did not rely on enhanced episodic item-memory for infrequent events for judgments on these measures, but rather on the more available evaluative information, which is in line with MCM predictions.

Despite the enhanced memory performance, the general IC pattern for both likability ratings and frequency estimates (apart from the influence of the missing D- behavior in study 1) remained similar across the 2 studies. This is in line with the MCM approach, which argues that these measures are based on the evaluative, rather than the episodic component of the information. In contrast, both the DBE and ILA argue that there should be a strong correlation between (true) memory performance and the extent of the illusion on these measures. Indeed, if we take the memory sensitivity measures from the multilevel measures (*BOLDItem*, which is similar to *d* from signal detection analysis - see the appendix) and relate it to global measures of frequency estimates and likability rating, we find no correlation between the interpersonal variation in the degree of the illusion in frequency estimates and impression ratings and memory performance. For the frequency estimates and the likability rating, the judgment for group A was subtracted from the judgment for group D for each participant. For both measures, higher values indicate more illusory correlation. As predicted by the MCM, neither frequency estimates, $r = -.22$ ($p > .16$), nor likability rating, $r = .03$ ($p > .82$), were significantly correlated with participants memory sensitivity, clearly supporting the MCM model.

General Discussion

Across two studies, we found a number of effects that are highly consistent with a multiple-component model of IC, which argues that participants base their judgments on the evaluative information embedded in a group prototype which is highly accessible, rather than episodic behavioral information, whenever possible. This was apparent in the almost linear relationship between group size and group impressions that emerged in both studies, and the finding that this pattern was independent of interpersonal variation in behavioral memory. Indeed, the introduction of an improved design and analysis, particularly in Study 2, allowed a more complete analysis of the

interplay between evaluative and episodic information. And the results from this study provided further support for the predictions of the MCM: As group size decreased, judgments based on item-memory revealed enhanced episodic memory for smaller groups and negative items, while judgments thought to be based on evaluative information showed the expected linear decrease in positivity.

Implications for illusory correlation accounts

In addition to the empirical support for the MCM, some of the results can be interpreted as being less than consistent with the MCM, and in fact appear to provide support for both distinctiveness-based and alternative accounts. In that sense, the findings from our studies parallel the IC literature as a whole. Although a number of findings provide clear support for the MCM model, the whole set of results does not allow to eliminate any of the alternative accounts as viable. This further strengthens the point already raised, namely that the phenomena associated with illusory correlation are likely to be determined by multiple processes and types of information. Nevertheless, our studies shed light on the interplay of episodic and evaluative information in producing illusory correlation, and present a number of results that have implications for the different accounts of, and further research into, illusory correlation.

Our studies demonstrated a number of findings that appear to be highly inconsistent with the original distinctiveness based explanation, which would predict that the illusion should appear only for the smallest and most undesirable group behaviors because these are most distinct. Our results show an increasing illusion from larger to smaller groups, when no distinctive behaviors are present (Study 1) *and* when D- items are in fact better encoded in memory (Study 2). These findings are problematic for the distinctiveness interpretation. Arguably, one could extend the DBE explanation to include more than one small group. Indeed, groups B and C could be considered to be more distinctive relative to Group A, and so one might expect some reduced IC effect for these groups as well (see Sherman et al., 1989). On the other hand, one could also argue that being of intermediate size and flanked by a larger and smaller group, groups B and C are in fact not very much salient or distinctive. The same issue arises when discussing the differentiation account and, to a lesser extent, the AT model.

Although these accounts can probably provide an explanation for our findings in terms of a meaningful, contextually-driven differentiation between the groups, it remains to be seen whether our results can be integrated into a wider process account of IC that avoids the inherent ambiguity of the distinctiveness concept (see also Fiedler, 2000).

Neither the Fiedler (1991) nor McGarty et al. (1993) models give any special status to the rare behaviors of a small group. These models have trouble accounting for the enhanced memory for infrequent information, as described above. Fiedler's model could possibly account for these results by making a number of auxiliary assumptions concerning stimulus salience, for instance by introducing a parameter for increased weighting or reduced noise for infrequent events (see Fiedler, 2000). However, the ILA would still predict that "the interpersonal variation in the degree of the illusion (in frequency estimates and impression ratings) is highly predictable from measures of memory performance" (Fiedler et al., p129, 1993), which was refuted by the findings of study 2. In general, Study 2 revealed item-memory effects that support the MCM and DBE. These results are inconsistent with previous studies that seemed to support the information-loss approach. However, they are quite consistent with a number of recent studies that strongly suggest that we might need to empirically revisit the distinctiveness approach, in order to explore its boundary conditions and the way enhanced item-memory is in fact causally related to IC (see for instance Risen, Gilovich & Dunning, 2007; Sherman, Conrey, & Groom, 2004; Sherman, Kruschke, Sherman, Percy, Petrocelli & Conrey, 2009). The current results do not settle the empirical argument, but strongly suggest that an explanation of illusory correlation effects purely in terms of aggregation and uniform information loss is insufficient to account for the full range of the current findings.

As mentioned in the introduction, Sherman and colleagues (2009) recently proposed an IC model based on Attention Theory (Kruschke, 2001, 2003). Both the MCM and AT model have a common ancestor in the Rescorla and Wagner model of associative learning (1972), but essentially represent different approaches to learning through error reduction (see Kruschke, 2001). Where the MCM is focused on the processing and representation of different types of information, models based on

attention theory describe how experience affects learning rates through attentional shifts. This allows the AT model to address shifts in learning due to different stimulus dimensions, which the MCM currently does not address explicitly. One could argue that the absence of differential reading times could be interpreted as being inconsistent with the AT model (Sherman et al., 2009). However, the current studies did not provide a direct test for the attention hypothesis, which would involve a design in which different stimulus dimensions such as common and rare attributes directly compete for attention, rather than being presented separately and sequentially. In fact, a design similar to the current one, but involving at least two distinct dimensions instead of a single global evaluative dimension (positive versus negative), would allow exploring the differences and similarities between the MCM and AT model. Successful attempts have been made to integrate both approaches (Kruschke, 1992). Applying such models to further explore how different types of information contribute to the IC phenomenon provides a critical avenue of future research.

What can the present results contribute to our understanding of the cognitive process behind IC? The results strongly suggest that any explanation of IC must distinguish between at least evaluative and episodic components (see also Johnson & Mullen, 1994; Johnson et al., 2001). A semi-distributed representation or multi-component analysis, in which different aspects of stimulus information are encoded at different levels, could provide an answer to contradictory results from past research which revealed evidence for enhanced memory for infrequent events in some studies (Hamilton & Sherman, 1989; McConnell et al., 1994) but not in others (Fiedler, 1991; 1996; Klauer and Meiser, 2000; Sanbonmatsu, Shavitt, Sherman & Roskos-Ewoldsen, 1987). These conflicting results might result from the way information in the IC task is represented: on the evaluative or episodic level. It is possible that in some studies, IC was tapped mainly at the evaluative level (e.g., Fiedler, 1991) while other studies directed the focus on memory performance at the episodic level (e.g., McConnell et al., 1994). Such a multi-component approach would be broadly consistent with dual-process models of group perception (Dovidio, Evans, & Tyler, 1986; Stangor, Sullivan, & Ford, 1991; Stephan, Ageyer, Coates-Shrider, Stephan, & Abalakina, 1994), attitudes (Edwards, 1990; Edwards & von Hippel, 1995) and person perception (Fiske & Neuberg, 1990; Jussim, Nelson, Manis, & Soffin, 1995). It

would also build on work that shows that perceivers can construct coherent prototypes from a wide range of information (Mayer & Bower, 1986) and is consistent with the idea that social information such as traits can be abstracted spontaneously during stimulus sentence processing (Uleman, Hon, Roman, and Moskowitz, 1996; Van Overwalle, Drenth, & Marsman, 1999). Importantly, an integrated process model of both evaluative and item-memory processes, as suggested by the MCM, would allow investigating systematically how the interaction between different processes contributes to the emergence of illusory correlations, and all of its associated outcomes.

Conclusion

The current studies confirm the predictions of a Multiple-Component Model about how memory performance would differ between various kinds of tasks in the illusory correlation paradigm, and under which circumstances these memory differences would be related to the perception of illusory correlations. First, it was shown that the typical IC bias as revealed in likeability ratings, frequency estimates and group assignments, appears to be driven by aggregate, evaluative information rather than differential memory for a certain stimulus category. This was clearly demonstrated by our group size manipulation in both studies, showing how the evaluative bias underlying the typical IC effects displayed a monotonic increase as group size decreases. Although a key prediction of both the MCM and ILA, the current studies are the first to produce this clear linear trend. The premise that contextual distinctiveness (McGarty et al, 2003; Sherman et al., 2009), or enhanced memory for distinctive undesirable minority behavior (Hamilton & Gifford, 1976; McConnell et al., 1994) is a necessary condition for the formation of IC was also refuted in both experiments, and especially in Study 1, where the illusion appeared for a group that contained no distinctive behaviors. The novel design in study 2 demonstrated enhanced memory for discrete episodic traces as group size decreased, and particularly for the smallest group – however, contrary to distinctiveness-driven explanations, this did not affect the general pattern of evaluative IC effects, which remained strong.

In conclusion, we argue that the whole set of results suggests that a distributed memory representation, in which evaluative and item-specific information is represented separately, can potentially provide a more complete understanding of IC. Although no single model can perfectly explain all the results, it is clear that any approach to IC will have to address the different types of information that underlie the phenomenon. Approaches that rely exclusively on either item-memory or regression principles can only provide partial explanations of the rich data set associated with the phenomenon, and will fail to explain how illusory correlation can occur both with and without enhanced memory, and how variations in the extent of the illusion are unrelated to variation in behavioral memory. Although the multi-component model (MCM) did not receive full support, it does provide a potential explanation for these apparently contradictory results.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed-effects models using Eigen and Eigen++. Version 0.999375-20. <http://cran.r-project.org/>
- Berndsen, M., & Spears, R. (1997). Reinterpreting illusory correlation: from biased covariation to meaningful categorisation. *Swiss Journal of Psychology*, *57*, 127-38.
- Berndsen, M., Spears, R., McGarty, C. & van der Pligt, J. (1998). Dynamics of Differentiation: Similarity as the Precursor and Product of Stereotype Formation. *Journal of Personality and Social Psychology*, *74*, 1451-1463.
- Chapman, L.J. & Chapman, J.P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, *72*, 193-204.
- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*, 335–359.
- Dawes, R.M. (1989). Experience and validity of clinical judgment: The illusory correlation. *Behavioral Sciences and the Law*, *7*, 457-467.
- DeCarlo, L. T. (1998). Signal Detection Theory and Generalized Linear Models. *Psychological Methods*, *3*, 186-205
- Dovidio, J. F., Evans, N., & Tyler, R. B. (1986). Racial stereotypes: The contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22- 37.
- Edwards, K. (1990). The interplay of affect and cognition in attitude formation and change. *Journal of Personality and Social Psychology*, *59*, 202-216.

- Edwards, K., & von Hippel, W. (1995). Hearts and minds: The priority of affective versus cognitive factors in person perception. *Personality and Social Psychology Bulletin*, 21, 996-1011.
- Fiedler, K. (1991). The tricky nature of skewed frequency tables: An Information loss account of distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, 60, 24-36.
- Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multi-cue environments. *Psychological Review*, 103, 193-214.
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107, 659-676.
- Fiedler, K., Kimmelmeier, M. & Freytag, P. (1999). Explaining asymmetric intergroup judgments through differential aggregation: Computer simulations and some new evidence. *European Review of Social Psychology*, 10, 1-40.
- Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory correlation and memory performance. *Journal of Experimental Social Psychology*, 29, 111-136.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, 23, pp. 1-74. New York: Academic Press.
- Hamilton, D.L. (1981). *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale, NJ: Erlbaum.
- Hamilton, D.L., & Gifford, R.K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, 12, 392-407.
- Hamilton, D.L., & Sherman, J.W. (1989). Illusory correlations: Implications for stereotype theory and research. In D. Bar-Tal, C.F. Graumann, A.W. Kruglanski & W. Stroebe (Eds.), *Stereotype and Prejudice: Changing conceptions* (pp. 55-92). New York: Springer.

- Hamilton, D.L., Dugan, P.M., & Trollier, T.K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlation. *Journal of Personality and Social Psychology, 48*, 5-17.
- Haslam, S. A., McGarty, C., & Brown, P. M. (1996). The search for differentiated meaning is a precursor to illusory correlation. *Personality and Social Psychology Bulletin, 22*, 611-619.
- Hintzman, D. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93*, 411-428.
- Jaeger, T. F. (2008). Categorical data analysis- Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434-446.
- Johnson, C. & Mullen, B. (1994). Evidence for the accessibility of paired distinctiveness in distinctiveness-based illusory correlation in stereotyping. *Personality and Social Psychology Bulletin, 20*, 65-70.
- Johnson, C., Mullen, B., Carlson, D. D., & Southwick, S. (2001). The affective and memorial components of distinctiveness-based illusory correlations. *British Journal of Social Psychology, 40*(3), 337-358.
- Jussim, L., Nelson, T. E., Manis, M., & Soffin, S. (1995). Prejudice, stereotypes, and labeling effects: Sources of bias in person perception. *Journal of Personality and Social Psychology, 68*, 228-246.
- Kelley, H. H. (1971). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.) *Attribution: Perceiving the causes of behavior* (pp. 1-26). Morristown, NJ: General Learning Press.
- Klauer, K., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review, 17*(4), 465-478.

- Klauer, K.C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, 26, 1074-1093.
- Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45(6), 812-863.
- Mayer, J. D., & Bower, G. H. (1986). Learning and memory for personality prototypes. *Journal of Personality and Social Psychology*, 51, 473-492.
- McClelland, J.L. & Rumelhart, D.E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs and exercises*. Cambridge, MA: Bradford.
- McConnell, A. R., Sherman, S.J., & Hamilton, D.L. (1997). Target entitativity: Implications for information processing about individual and group targets. *Journal of Personality and Social Psychology*, 72, 750-762.
- McConnell, A.R., Sherman, S.J., & Hamilton, D.L. (1994). Illusory correlation in the perception of groups: An extension of the distinctiveness-based account. *Journal of Personality and Social Psychology*, 67, 414-429.
- McGarty, C., Haslam, S. A., Turner, J. C., & Oakes, P. J. (1993). Illusory correlation as accentuation of actual intercategory difference: Evidence for the effect with minimal stimulus information. *European Journal of Social Psychology*, 23, 391-410.
- Murphy, R. A., Schmeer, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear- The effects of increased learning. *Quarterly Journal of Experimental Psychology*, 64(1), 24-40.
- Pryor, J. B. (1986). The influence of different encoding sets upon the formation of illusory correlations and group impressions. *Personality and Social Psychology Bulletin*, 12, 216-226.
- Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.) *Classical conditioning II: Current research and theory* (pp. 64–98). New York: Appleton-Century-Crofts.

- Sanbonmatsu, D.M., Shavitt, S., Sherman, S.J. & Roskos-Ewoldsen, D.R. (1987) Illusory correlation in the perception of performance of self or a salient other. *Journal of Experimental Social Psychology*, 23, 518-543.
- Seligman, M.E.P. (1975). *Helplessness: On depression, development and death*. San Francisco, CA: Freeman.
- Shavitt, S., Sanbonmatsu, D. M., Smittipatana, S. & Posavac, S. S. (1999). Broadening the conditions for illusory correlation formation: Implications for judging minority groups. *Basic and Applied Social Psychology*, 21, 263—279.
- Sherman, S.J., Hamilton, D. L. & Roskos-Ewoldsen, D. R. (1989). Attenuation of Illusory Correlation. *Personality and Social Psychology Bulletin*, 15, 559-571
- Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory and Cognition*, 33(1), 151-170.
- Slotnick, S. D., Dodson, C. S., Klein, S. A., & Shimamura, A. P. (2000). An Analysis of Signal Detection and Threshold Models of Source Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 26(6), 1499-1517.
- Smith, E.R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology*, 27, 107-123.
- Smith, E. R. & DeCoster, J. (1998). Knowledge acquisition, accessibility, and use in person perception and stereotyping: Simulation with a recurrent connectionist network. *Journal of Personality and Social Psychology*, 74, 21—35.
- Stangor, C., Sullivan, L. A., & Ford, T. E. (1991). Affective and cognitive determinants of prejudice. *Social Cognition*, 9, 359-380.
- Stephan, W. G., Ageyev, V., Coates-Shrider, L., Stephan, C. W., & Abalakina, M. (1994). On the relationship between stereotypes and prejudice: An international study. *Personality and Social Psychology Bulletin*, 20, 277-284.

- Stroessner, S.J., Hamilton, D.L., & Mackie, D.M. (1992). Affect and stereotyping: The effect of induced mood on distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology*, 62, 564-576.
- Swets, J.A. (1973). The relative operating characteristic in psychology. *Science*, 182, 990-1000.
- Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, 22, 377-394.
- Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or the action? *Personality and Social Psychology Bulletin*, 25, 450-462.
- Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review*, 110, 536-563.
- Vonk R. 1998. The slime effect: Suspicion and dislike of likeable behavior toward superiors. *Journal of Personality and Social Psychology*, 74, 849-864.
- Wright, D. B., & London, K. (2009). Multilevel modeling: Beyond the basic applications. *British Journal of Mathematical & Statistical Psychology*, 62, 439-456.
- Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*, 41, 257-267.

Table 1

Number of statements presented as a function of group (A, B, C or D) and type of behavior (desirable versus undesirable) in Experiment 1 and 2.

Group:	A	B	C	D
Desirable behaviors	16	8	4	2
Undesirable behaviors	8	4	2	0 / 1 ^a

^a In Experiment 1, group D contained no undesirable behaviors while in Experiment 2, group D contained 1 undesirable behavior.

Table 2

Experiment 1: Group Impression and Behavioral Memory as a function of type of behavior and group.

	Group				Group Size F(3, 222) ^a	Linear Trend F(1,74)
	A	B	C	D		
Likeability Rating (Mean)	5.37	4.72	4.56	4.36	5.15**	8.82**
Frequency Estimation (Mean)	.38	.47	.47	.38	3.48*	ns
Group Assignment (Observed and Actual Proportion ^b)						
Desirable	.48 (.53)	.24 (.27)	.17 (.13)	.08 (.07)		
Undesirable	.36 (.57)	.29 (.28)	.24 (.14)	.07 (.00)	51.88***	n/a
Free Recall (Proportion Correct)						
Desirable	.16	.14	.14	.22	2.50°	ns
Undesirable	.28	.25	.25	—	ns	ns

° $p < .06$, * $p < .05$, ** $p < .01$, *** $p < .001$

^a The degrees of freedom may actually be smaller in some cases as some data points were missing. The effect of group size was analyzed across desirable and undesirable behaviors, except for the memory measures (see text).

^b Proportion = Number of assigned (un)desirable behaviors divided by total number of (un)desirable behaviors.

Table 3

Experiment 2: Group Impression and Behavioral Memory in function of type of behavior and group.

	Group				Group Size F(3, 126) ^a	Linear Trend F(1,42)
	A	B	C	D		
Likeability Rating (Mean)	6.87	5.70	5.66	5.22	9.91***	13.87***
Frequency Estimation (Mean)	.32	.46	.50	.57	4.94**	7.95**
Group Assignment (Observed and Actual Proportion ^b)						
Desirable	.34 (.53)	.15 (.27)	.10 (.13)	.06 (.07)		
Undesirable	.14 (.53)	.12 (.27)	.09 (.13)	.06 (.07)	87.71***	n/a
Free Recall (Proportion Correct)						
Desirable	.25	.19	.20	.29		
Undesirable	.19	.22	.21	.33	3.87**	5.39*

* $p < .05$, ** $p < .01$, *** $p < .001$ ^a The effect of group size was analyzed across desirable and undesirable behaviors (see text).^b Proportion = Number of assigned (un)desirable behaviors divided by total number of (un)desirable behaviors.

Appendix

Accuracy analysis

The present analysis uses accuracy, i.e. participants' ability to correctly assign an item to a group, as dependent variable. The logit is a popular measure of accuracy: For the current analysis, it corresponds to the natural logarithm of the ratio of the odds of assigning an item to group X when the item is from Group X, over the odds of assigning an item to group X when the item is not from Group X. Zero corresponds to chance responding, while values above zero show memory. The population logit can be estimated by the coefficient in a multilevel logistic regression (Wright, Horry, & Skagerberg, 2009). A number of computer programs allow random effects to be modelled with the following multilevel logistic regression (Baayen, Davidson, & Bates, 2008; Wright & London, 2009):

$$\text{logit}(p[\text{Accuracy}_{ij}]) = \beta_0_j + \beta_1 \text{Linear}_j$$

Let Accuracy_{ij} be the accuracy on i th trial for the j th person, and β_0_j is the intercept. The subscript j means it has different values for each participant. We assume there is some grand mean, β_0 , and that values for participants will be normally distributed around this value with some unknown standard deviation to be estimated. $\beta_1 \text{Linear}_j$ is a vector coding for a linear trend. The subscript j means that the coefficient value can vary by participant. We assume that these values are normally distributed around β_1 . Multilevel generalized linear models are currently fit with methods that approximate maximum likelihood. For the current analysis, we used the Laplace method, which is the default for the package lme4 (Bates, Maechler, & Dai, 2008). It is part of the software R (R Development Core Team, 2008) and is available for free from the Comprehensive R Archive Network.

Decoy analysis

For this analysis, we created a separate binary DV by recoding participants' responses in such a way that an assignment to groups A – D corresponded to an "old" answer (coded 1), while a "decoy" answer corresponded to a "new" answer (coded 0). For any individual trial, a participant thus either categorizes an item as 'old' (item was presented before) or 'new' (item is a decoy). The probability of saying 'old' should increase if the item was presented previously, provided performance is above

chance. The probability of an 'old' response will vary among participants, and we also expect there to be differences among participants for how good their memories are. Therefore, we allow random effects for participants (see also Clark, 1973). This can be modelled with the following multilevel logistic regression (Baayen, Davidson, & Bates, 2008; Wright & London, 2009):

$$\text{Logit}(\text{Pr}[\text{SaysOld}_{ij}]) = \beta_0_j + \beta_1 \text{OldItem}_j$$

Let response_{ij} be the response on *i*th trial for the *j*th person. The β_0_j is the intercept and relates to the response criterion in signal detection theory terminology (Wright & London, 2009). The subscript *j* means it has different values for each participant, which in SDT terminology corresponds to people having different response criteria. We assume there is some grand mean, β_0 , and that values for participants will be normally distributed around this value with some unknown standard deviation to be estimated. $\beta_1 \text{OldItem}_j$ provides a measure of accuracy. The subscript *j* means that values can vary by participant. Again, we assume that these values are normally distributed around β_1 . Participants performance should be above chance, so we expect that $\beta_1 > 0$. The coefficient for whether an item was seen before or not (here OldItem) measures discriminability (a measure of memory sensitivity comparable to *d'*, see De Carlo, 1998) and interactions between this and other variables show whether these other variables moderate accuracy (London & Wright, 2009).

Of particular interest here is whether accuracy is moderated by group size or valence. In other words, we want to test whether accuracy (measured by OldItem) depends on whether an item is from a particular group, or is positive or negative in sign. This can be explored by testing interaction effects between β_1 and other variables. This corresponds to moderator analysis, where one explores if the effect of one variable depends on the level of another variable. We used a stepwise approach to search for moderator variables. The main effect of each added component and its interaction with OldItem were included in the model. For instance, to test if valence of the item (negative versus positive) moderates accuracy, we first include the main effect of valence. If this improves the fit of the model, we subsequently tested the interaction between this term and β_1 . The same approach was used for all other variables mentioned in the text. As with the accuracy model, we used the Laplace method to fit the models.