British Journal of Social Psychology (2001), **40,** 209–233 *Printed in Great Britain*

Meaning-seeking in the illusory correlation paradigm: The active role of participants in the categorization process

Mariëtte Berndsen*

University of Amsterdam, The Netherlands

Craig McGarty

The Australian National University, Australia

Joop van der Pligt and **Russell Spears**

University of Amsterdam, The Netherlands

The present research examines the role of categorical perception (McGarty, Haslam, Turner, & Oakes, 1993) in the illusory correlation paradigm. This approach assumes that the search for meaningful differences between two stimulus groups can lead to the illusory correlation effect. This explanation is investigated in Study 1 by presenting participants with constrained stimulus information and examining whether accentuating evaluative differences between stimuli could provide a basis for illusory correlation. Results of this study (N = 64) revealed illusory correlation effects that were related to evaluative reinterpretations of the stimuli. Study 2 (N = 19) focused on the causal relation between illusory correlation and accentuation effects, using the thinking-aloud technique. Detailed analyses of each participants' behaviour indicated that illusory correlation can arise from actively seeking intergroup differences and that reinterpretations of stimuli and the perception of illusory correlation were mutually reinforcing. Implications of these results for stereotyping processes are discussed.

'I'm sure I didn't mean—' Alice was beginning, but the Red Queen interrupted her impatiently. 'That's just what I complain of! You *should* have meant! What do you suppose is the use of a child without any meaning? Even a joke should have some meaning—'

L. Carroll, 1960, p. 319

In order to make sense of the world, we often go beyond the given information. However, in this inferential process, we sometimes arrive at erroneous conclusions or judgments. One of these 'incorrect' judgments is called 'illusory correlation'. That is, people sometimes perceive an erroneous covariation between two sets of events. These illusory correlations have been widely researched in social

* Requests for reprints should be addressed to Mariëtte Berndsen, Department of Social Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: sp_berndsen@macmail.psy.uva.nl).

psychology, partly because this mechanism could offer an explanation for the acquisition of stereotypes about minorities. In the present study we investigated the processes which produce the illusory correlation effect. More specifically, we investigated the possibility that illusory correlation arises as a product of categorical perception (McGarty, Haslam, Turner, & Oakes, 1993).

The term 'illusory correlation' refers to the perception of covariation between two classes of stimuli which are uncorrelated or less strongly correlated than perceived. Hamilton and Gifford (1976) developed a now familiar paradigm to demonstrate the illusory correlation effect. They showed that the combination of a (statistically) infrequent behaviour and an infrequently encountered group was overestimated, producing what is termed a 'distinctiveness-based illusory correlation'. According to Hamilton and Gifford, statistically infrequent categories are particularly distinctive for the perceiver so that they receive more attention, are more efficiently encoded, and consequently are more available in memory than less distinctive categories. Although this explanation has been questioned by a number of researchers (Fiedler, 1991; McGarty *et al.*, 1993; Smith, 1991), there is also considerable support for distinctiveness-based illusory correlation (see e.g. McConnell, Sherman, & Hamilton, 1994; Mullen & Johnson, 1990). Moreover, the approaches of Hamilton, Smith, and Fiedler all focus on the perception of infrequent information as the critical variable thought responsible for the effect.

In contrast, research by McGarty *et al.* (1993) has shown that the perception of infrequent stimuli is not necessary to obtain illusory correlations. They argued that illusory correlation could result from a categorical differentiation process. Their explanation is based on self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) which in turn is rooted in Tajfel's work (e.g. Tajfel, 1969). According to these accounts, categories are formed on the basis of the perception of similarities among and differences between stimuli. Categorization effects reflect the tendency to enhance the differences between distinct categories on a certain dimension and to minimize the differences within each of the categories. Comparative fit (Oakes, Turner, & Haslam, 1991) is enhanced when the differences between categories are larger than differences within the categories. However, in order to accentuate differences between Groups A and B, it is necessary to perceive differences between them. According to McGarty *et al.*, these differences exist in the classic task of Hamilton and Gifford (1976, Study 1) because there is more evidence for the hypothesis that Group A is 'good' and Group B is 'bad' (18 + 4 = 22 stimuli) than for the opposite hypothesis (9 + 8 = 17 stimuli). The positivity of Group A over Group B thus reflects the comparative fit between group membership and behaviour, and forms the basis for further accentuating the differences between the groups.

Following this line of reasoning and drawing on the concept of 'differentiated meaning' (McGarty & Turner, 1992), McGarty *et al.* (1993) argued that participants in the illusory correlation paradigm try to make sense of the stimulus situation by perceiving the stimuli in terms of meaningful (i.e. clear and separable) categories. Given the content of the stimuli presented, meaningful categories to differentiate between the stimulus groups are provided by positive and negative behaviours. McGarty *et al.* proposed that this sense-making process occurs because the task

situation creates expectations of intergroup differences on the behavioural dimension. This can be related to research on the effects of the conversational context (Bless, Strack, & Schwartz, 1993; Gigerenzer, 1991; Grice, 1975; Hilton, 1995; Schwarz & Bless, 1992; Stapel, Reicher, & Spears, 1995), which can also be applied to settings in which the experimenter communicates information to participants. Participants expect that the information presented in a task is informative, meaningful, true and relevant. Bless *et al.* showed in different research domains that it is possible that participants' inferred meaning differs from researcher's intended meaning. They argued that participants try to determine the meaning of the task not because they want to behave as a 'good participant' (Orne, 1962), but rather because they follow the conversational rules of being informative, relevant, truthful and unambiguous (Grice, 1975). Abele and Petzold (1998) showed that participants try to use all information presented by making sense of both explicit information and subtle cues ('meta-information') in the task.

Applying the previously mentioned conversational rules to the illusory correlation task, participants are likely to ask themselves: 'Why would the experimenter ask me whether the groups are different unless they were?' In other words, they might expect some *meaningful* difference between the groups because they have been given different names (McGarty & Turner, 1992). Accordingly, the illusory correlation effect may partly be based on participants' assumption that there must be a meaningful difference between the groups. Indirect support for this view is provided by Haslam, McGarty, and Brown (1996) and Berndsen, Spears, Van der Pligt, and McGarty (1999).

Haslam *et al.* (1996) and Berndsen *et al.* (1999) provided evidence that when distinctiveness is present but there is no reason to search for differences between the groups, the illusory correlation effect disappears. The other side of this coin is to ask whether the illusory correlation effect does occur when there are reasons to differentiate between the stimulus groups but when there is no distinctiveness. The aim of the present study is to investigate this idea and to extend the findings of McGarty *et al.* (1993).

One possible way to investigate the differentiated meaning perspective in the absence of distinctiveness is by *constraining* the stimulus information. In the standard illusory correlation task, participants are presented with stimuli that vary on a behavioural dimension (positive and negative behaviours) and a group dimension (two groups, called Group A and Group B). In the present study we constrained the type of stimulus information by presenting either one type of behaviour emanating from two groups, or two types of behaviour displayed by members of one group only. Thus, in Study 1, we tested the effects of constrained stimulus information on the perception of illusory correlation. If participants were indeed able to find a basis for differentiating between the stimulus groups, we also had to investigate how they did this. McConnell *et al.* (1994) suggested that the perception of illusory correlation can involve an interpretive process, where the meaning of old information is changed in the light of new information and experiences. This suggestion can be investigated by using process measures that could provide insight into whether the information about the meaning of the stimuli is changed by its context. Recent research on this topic (Berndsen, Van der

Pligt, Spears, & McGarty, 1996; Berndsen, Spears, McGarty, & Van der Pligt, 1998) showed that participants reinterpret the stimuli in the light of their categorical perception of them. Berndsen and colleagues used a rating task in which participants were presented with behavioural statements ascribed to Group A and B, and were asked to rate these statements on a desirability scale. There was no correlation between type of behaviour and group membership, and Group B's negative behaviours were distinctive in terms of their statistical infrequency (as in the traditional paradigm). As predicted, illusory correlation effects occurred when participants judged the statements in terms of their valence rather than just perceiving them. Moreover, the response pattern on the rating task revealed that participants started to differentiate between the groups in the second half of that task as indicated by more polarized ratings. In the present study we also used this rating task in order to examine the process underlying the formation of illusory correlation.

Fiedler, Hemmeter, and Hofmann (1984) demonstrated that stimuli are subjected to differential interpretations. In their study this was because of existing opinions about the stimulus groups. Their participants were asked to rate each attitude statement ('liberal vs. authoritarian', produced by students or clerks) on an authoritarian–liberal dimension during stimulus presentation. Results showed that statements were interpreted in line with prior opinions, or expectations, indicating that the 'illusion' is already effective during the perception (and evaluation) of the stimuli. Unlike this expectation-based research, there is, as far as we know, no previous study of data-based illusory correlation that attempted to measure impression-formation processes without the possible confound of distinctiveness. That is, Berndsen *et al.*, (1996, 1998) used process measures but in the presence of distinctive stimulus information. The purpose of Study 1 was to test whether and how participants differentiate between the groups in the absence of distinctive information. As was found by Berndsen *et al.* (1996, 1998), we expected that *evaluative reinterpretations* of the statements would be related to the illusory correlation effect.

STUDY 1

According to McGarty and colleagues (1993; McGarty & de la Haye, 1997), features of the standard task, such as the presence of two groups and a comparative dimension of evaluation, are the basic elements in producing illusory correlations, and not stimulus distinctiveness. In order to investigate this differentiated meaning account in the present study, we avoided paired-distinctive stimulus information about the groups. This was achieved by constraining the stimulus situation by either presenting information about only one stimulus group or by presenting only positive information. If the illusory correlation effect depends upon finding a basis to differentiate between groups rather than distinctiveness, it follows that the effect should also have occurred in the present experiment.

We created two sets of constrained tables. The traditional 2×2 contingency table of the illusory correlation paradigm involves a behavioural dimension and a

group dimension. We manipulated these dimensions in the constrained stimulus conditions. Participants in one condition were presented with positive behaviours performed by members of Group A (majority group) and Group B (minority group). This condition is termed 'behaviour-constrained condition' because in contrast to the traditional 2×2 contingency table, stimulus information on the behaviour dimension is constrained. Participants in the second condition were presented with positive (majority) and negative (minority) behaviours performed by members of Group A only. This condition is termed 'group-constrained condition' because stimulus information on the group dimension is constrained. Our first prediction was that participants in both conditions would display illusory correlation. This prediction is based on the fit between the two majority categories: there is more evidence for desirable Group A behaviours than for desirable Group B behaviours (behaviour-constrained condition). Likewise, in the group-constrained condition there is more evidence for desirable than for undesirable Group A behaviours.

The present study also investigated whether participants in the constrained stimulus conditions would reinterpret their evaluations of the various behaviours. Therefore, we asked participants afterwards to rate each of the perceived statements on an evaluative dimension. These evaluations were compared with those of a control condition in which there was no illusory correlation task (and participants were simply asked to rate each statement without references to any groups). Our second prediction was that participants in the constrained stimulus conditions would reinterpret the perceived behaviours in order to maximize the evaluative differences between groups. That is, participants in the behaviour-constrained condition would downgrade the valence of the positive behaviour performed by Group B compared to the control condition (prediction 2a). A similar prediction (prediction 2b) concerned participants in the group-constrained condition who were expected to upgrade the negative behaviours performed by Group A by making them less positive (again compared to the control condition). Finally, our third prediction was that illusory correlation would be related to accentuated ratings of the statements. We predicted that participants in the behaviour-constrained condition who displayed high illusory correlations should also have shown high differences in the evaluative ratings of the positive statements performed by Groups A and B. With respect to the group-constrained condition, we predicted that participants who displayed high illusory correlations should also have shown small differences in the evaluative ratings of the positive and negative statements performed by Group A, because participants would evaluate the negative Group A behaviours as less negative.

Method

Participants and design

Participants were 64 first-year psychology students at the Australian National University who participated in the study as part of practical classes. Participants were assigned randomly to one of the

three conditions: the behaviour-constrained condition, the group-constrained condition, and the control condition.

Stimulus materials and procedure

In a pilot study, 93 behaviour descriptions were rated by 25 participants on 9-point scales ranging from 1 = very undesirable to 9 = very desirable. A selection of these descriptions was used in the experiments: 36 items were moderately positive and 12 items were moderately negative.

Using similar instructions to those of Hamilton and Gifford (1976), participants were informed that they would be shown behavioural descriptions about members of two groups, labelled A and B, and that in the real world Group B is smaller than Group A. Consequently, they were told that statements describing members of Group B would occur less frequently. Participants were also told that they would be presented with a *selection* of all the statements, because it would take too much time to read them all. Next, participants in the behaviour constrained-condition were presented with 24 positive statements about members of Group A and 12 positive statements about members of Group B. Participants in the group-constrained condition were presented with 24 positive and 12 negative behaviours performed by members of Group A. After viewing the statements, participants were asked to complete four dependent measures.

Dependent measures

For the trait rating task, all participants were asked to evaluate Groups A and B on 9-point rating scales: 1 = unpleasant to 9 = pleasant; 1 = friendly to 9 = unfriendly, 1 = unsympathetic to 9 = sympathetic and 1 = reliable to 9 = unreliable. For the frequency estimation task, participants in the behaviour-constrained condition were informed that 36 positive behaviours were performed by members of Groups A and B and that there were also 18 negative behaviours performed by both groups which were not presented. Participants in the group-constrained condition were informed that 36 behaviours were performed by members of Group A and that there were 18 Group B members, whose behavioural descriptions were not shown. The participants were asked to estimate how many of the statements about members of both groups described desirable and undesirable behaviours.

The next task that participants had to complete differed from the traditional assignment task in which participants again received all the statements they had been presented with before, but without information about group membership. In standard experiments they are asked to indicate the group membership of the person who performed each of the behaviours. In the present experiments such a task could not be used because participants in the group-constrained condition, for example, would know already that all the perceived behaviours emanated from Group A. We therefore developed a different 'projective' assignment task in which all participants were exposed to 18 new statements of which 12 described positive behaviours and six described negative behaviours (see also Berndsen *et al.*, 1996). The participants were instructed to guess from which group each statement emanated.

The final task was the rating task in which participants were asked to rate each of the presented statements on a 9-point scale ranging from 1 = very undesirable to 9 = very desirable. Participants in the control condition were asked to rate all 48 statements perceived by participants in the constrained stimulus conditions (12 negative and 36 positive behaviours) on the same scale. These statements differed from those in the experimental conditions in that the group label was omitted. Thus, each statement described a particular behaviour without referring to group membership.

Results

Trait ratings

With respect to the perception of illusory correlation, we predicted that members of Group A would be evaluated more positively than members of Group B in the

Table 1. Illusory correlation as a function of constrained stimulus information (Study 1)

	Behaviour-constrained condition	Group-constrained condition
Trait rating		
Mean Group A	29.75	25.68
Mean Group B	27.17	20.96
Difference	2.58*	4.73**
Estimation		
Positive A	22.42 (24)	24.70 (24)
Negative B	8.33 (6)	7.65 (6)
*Z*phi	.09*	.12**
New assignment		
Positive A	7.17 ($\geq$ 8)	7.70 ($\geq$ 8)
Negative B	3.21 ($\leq$ 2)	3.26 ($\leq$ 2)
*Z*phi	.14*	.19*

*$p < .01$; **$p < .001$.
Note: Level at which mean is different from zero (based on one-tailed *t* tests).

stimulus constrained conditions. After recoding the rating scales for friendly–unfriendly and reliable–unreliable, the four rating scales yielded a Cronbach's alpha of .85 (Group A) and .86 (Group B). For each group, one evaluative index-score was computed based on the four scales of the trait rating task. The possible range was 4 to 36; a higher rating indicates a more positive evaluation.

The mean ratings in Table 1 show that Group A was evaluated more positively than Group B. This difference was significant in both conditions.

Frequency estimates and assignments of behaviours to group

The results on the frequency estimate task and assignment task for the constrained conditions are also presented in Table 1. For both tasks the expected values concerning the illusory correlation effect are given in parentheses (the derivation of these expected values is presented in the Appendix). The phi and Fisher's *Z* distributions were generally normal (Haslam & McGarty, 1994) and *Z* scores were used.

In order to test our first prediction, we compared the mean transformed phi scores to zero. Results revealed significant differences on both tasks for both conditions. These findings correspond with those of the trait ratings.[1]

Evaluative ratings of the stimuli

Our second prediction was that participants in the constrained conditions would reinterpret the evaluations of the statements in order to create evaluative

[1]These findings replicate those of a previous experiment with the same constrained conditions and a standard condition in which participants were presented with 16 positive and eight negative behaviours performed by Group A, and eight positive and four negative behaviours performed by Group B. As predicted, illusory correlations occurred on the three measures for all conditions (all p's $< .01$).

Table 2. Mean evaluative ratings of the presented statements for all conditions (Study 1)

Evaluative rating task	Behaviour-constrained	Control	Group-constrained
Positive behaviours Group A	7.14	7.32	7.41
Positive behaviours Group B	6.64	7.43	–
Negative behaviours Group A	–	2.26	2.71

Note: Scale is from 1 to 9, higher means indicate more positive evaluations.

differences between the groups. Table 2 reports the mean evaluative ratings of the perceived statements for the constrained conditions and the ratings of these statements in the control condition.

Prediction 2a was that participants in the behaviour-constrained condition would downgrade the positive behaviours performed by Group B as compared to the control condition (based on the assumption that participants have associated Group A with positive behaviour and Group B with negative behaviour on the illusory correlation tasks). We excluded three participants with a so-called 'reversed illusory correlation'.[2] From Table 2 it can be seen that Group B was evaluated less positively in the behaviour-constrained condition than in the control condition. The interaction between these conditions and group (within-participants factor) was significant ($F(1,36) = 22.98$, $p < .001$). Prediction 2b was that participants in the group-constrained condition would upgrade the negative Group A behaviours compared to the control condition. This prediction was given some (marginally significant) support ($F(1,35) = 3.31$, $p < .08$).[3]

We also investigated whether the evaluative reinterpretations of the statements would be larger in the behaviour-constrained condition than in the group-constrained condition. We anticipated this effect because the relevant intergroup comparative context (behaviours performed by both Group A and Group B) is provided in the behaviour-constrained but not in the group-constrained condition. To be more specific, evaluating the negative Group A behaviours as relatively positive, or less negative (group-constrained condition), will increase the overall intragroup similarity in Group A, but does not provide an explicit basis for differentiating between Groups A and B. In contrast, evaluating the positive Group B behaviours as relatively negative (behaviour-constrained condition) produces an explicit intergroup contrast with Group A. The 'evidence' of unfavourable Group B behaviours in this condition should allow participants to derive clear differences between the categories in terms of Group A being better than Group B. Thus, we expected that statements would be more strongly reinterpreted in the behaviour-constrained condition than in the group-constrained condition. In order to test this hypothesis, we computed a difference score index for the evaluative ratings between

[2]Including the three participants with a reversed illusory correlation did not affect the results.
[3]Including the three participants with a reversed illusory correlation resulted also in a marginally significant interaction.

each constrained condition and the control condition. With respect to the behaviour-constrained condition we computed the overall difference in the Group B positive behaviours by subtracting from each participant's score the mean rating of the control group. Likewise, for the group-constrained condition we computed the overall difference in the Group A negative behaviours by subtracting from each participant's score the mean rating of the control group. Our prediction that this difference score-index would be greater in the behaviour-constrained ($M = .79$) than in the group-constrained ($M = .45$) condition was supported ($t(39) = 1.74$, $p < .05$, one-tailed).[4]

Our third prediction was that illusory correlation would be related to accentuation effects in the ratings of the statements. For the behaviour-constrained condition we predicted that participants who showed strong illusory correlations would also show strong differences between the ratings of Group A positive behaviours and Group B positive behaviours. After computing a difference score between the ratings of positive A and B behaviours, we correlated this difference score with the scores on the three illusory correlation measures. For the trait rating task we also computed the difference between the evaluation score for Group A and Group B. Our prediction was supported by a significant correlation between the evaluative ratings and the scores on the three tasks (trait ratings: $r = .44$, $p < .05$; frequency estimates: $r = .52$, $p < .01$; assignments: $r = .61$, $p < .01$).

Our third prediction suggests that participants in the group-constrained condition who displayed strong illusory correlations should also display small differences between the ratings of positive and negative Group A behaviours. In this case it was necessary to take participants with a 'negative illusory correlation' into account; these participants associated Group A with negative behaviour and Group B with positive behaviour. We reasoned that participants with both positive and negative illusory correlations would minimize the evaluative difference between the positive and negative behaviours performed by Group A in contrast to participants who did not display illusory correlation. Illusory correlation scores range from negative via zero to positive, indicating that the no-illusory correlation group is positioned in between the negative and positive illusory correlation groups. Because we expected that the middle group would see more evaluative differences within Group A than both other groups, it follows that the relationship between the evaluative ratings and the illusory correlation scores should be curvilinear rather than linear. Both kinds of relationships were tested in a regression analysis and the results are reported in Table 3.

From Table 3 it can be seen that there was a negative relationship between the evaluative ratings and the three measures of illusory correlation. This supported the prediction that the strong illusory correlations should be related to weak differences between the positive and negative Group A behaviours. Moreover, the addition of the curvilinear predictor to the linear predictor resulted in a significant increase in the explained variance of the evaluative ratings on both the frequency estimates and assignments tasks. The fact that there was no curvilinear, but rather a linear, relationship between the trait ratings and the evaluative ratings is not surprising

[4]Including the participants with a reversed illusory correlation resulted also in a significance difference.

Table 3. Hierarchical regression analysis of the evaluative ratings on measures of illusory correlation in the group-constrained condition (Study 1)

	Evaluative ratings			
	Linear		Curvilinear	
Predictor	β	R^2	β	R^2 change
Trait ratings	−.50	.25**	−.21	.04
Frequency estimates	−.34	.12	−.50	.22**
Assignments	−.08	.01	−.66	.17*

*$p = .05$; **$p < .05$.

because there was only one participant who evaluated Group A slightly negatively on the trait ratings. In sum, our prediction that illusory correlation is related to accentuation effects in the ratings of the perceived statements has been supported.

Discussion

This study demonstrates that incomplete information about groups and their behaviours can produce biased covariation judgments. Perceiving more positive than negative behaviours in a single group (group-constrained condition) seems sufficient to decide that the other group is relatively bad, although participants were *not* provided with information about the behaviour of that group. This is in accordance with findings of Acorn, Hamilton, and Sherman (1988), who demonstrated that judgments can generalize to behaviour that is not perceived. Likewise, the perception of more positive behaviours in Group A than Group B (behaviour-constrained condition), without knowing anything about the distribution of the negative behaviours, resulted in the view that Group A is relatively good and Group B is relatively bad. Thus results support the categorization approach of McGarty *et al.* (1993) and show that participants found a basis to differentiate between the groups.

The fact that illusory correlations can occur without introducing distinctiveness is problematic for the distinctiveness account (Hamilton & Gifford, 1976) or the extended distinctiveness account (McConnell *et al.,* 1994) of illusory correlation. These accounts predict no illusory correlation effects in the stimulus constrained conditions because of the absence of a paired-distinctive category. Although our results do not invalidate the distinctiveness account of Hamilton, and the models of Smith (1991) and Fiedler, Russer, and Gramm (1993) that focus on the role of memory, at best, these approaches would have to introduce additional parameters to account for these results.

Our results support the view that representations of the perceived stimuli are not static and that the reinterpretations of the stimuli correspond with the illusory

correlation effects. This is in accordance with the findings of Berndsen *et al.* (1996, 1998). Reinterpreting the stimuli in the light of categorical perception supports McGarty *et al.*'s (1993) notion of categorical perception. The present study extends this idea by showing *how* participants try to differentiate between groups. That is, participants in the behaviour-constrained condition interpret the positive behaviours performed by Group B as relatively negative in order meaningfully to differentiate between the groups. In the group-constrained condition, participants interpret the negative Group A behaviours as relatively positive as compared to the control condition, which allows them to view Group A members as positive.

Overall, participants in the behaviour-constrained condition reinterpret the evaluations of the behaviours more than participants in the group-constrained condition. This could well be caused by the fact that downgrading the positive behaviours performed by Group B allows these participants to differentiate *explicitly* between the two groups. In contrast, participants in the group-constrained condition could only produce *implicit* differential meaning between the groups by upgrading the negative Group A behaviours, resulting in more similarity in this group in the absence of information about the behaviours in Group B.

Alternatively, one might argue that learning mechanisms are responsible for the present findings. That is, seeing more positive behaviours performed by Group A than by Group B (behaviour-constrained condition) results in a more positive evaluation of Group A. Likewise, learning that Group A is mostly positive and knowing nothing about Group B (group-constrained condition) creates an advantage for Group A. However, this learning explanation has some shortcomings. First, if one learns that Group A is generally positive (in both constrained conditions), why should one judge Group B as relatively negative? If one has not been able to learn about this group, it would also be possible, for example, to evaluate Group B as positive such that the groups do not differ. Therefore, the consistent finding that one views Group B as relatively bad cannot be explained by learning mechanisms, and supports the intergroup differentiation approach of McGarty *et al.* (1993). Most importantly, the present results show that reinterpretive processes (rather than purely learning mechanisms) can give rise to intergroup differentiation.

One issue which this experiment cannot resolve is the *direction of the causal relation* between illusory correlation and the accentuation effects in the evaluative judgments of the perceived behaviours. Given the order of the tasks in the present study, it might be that the perception of illusory correlation subsequently affected the meaning of the statements. Although Berndsen *et al.* (1996, 1998) showed that reinterpretations also occurred in the rating task *before* completing the illusory correlation tasks, they did not show whether the process of reinterpretation causes the illusory correlation effects or vice versa. With respect to the latter option, one could distinguish two stages. The first consists of inferring a fit between group membership and desirability of the behaviours in the illusory correlation task. The second stage is that once participants have formed a (biased) impression of the groups, this impression will subsequently be confirmed in the ratings of the perceived statements. We expect that the processes in these two stages are not

totally separated and influence each other. In Study 2 we investigated whether both processes (i.e. forming impressions and reinterpreting) influence each other by asking participants to think aloud during the experimental session. The benefit of the thinking-aloud method is that it allowed us to examine processes underlying the perception of illusory correlation.

STUDY 2

The present study investigates whether the behaviours in the standard task are subjected to differential interpretations. We used the traditional task as introduced by Hamilton & Gifford (1976), the only difference being that we asked participants to think aloud during the whole experimental session. The purpose of the thinking-aloud method was twofold: (1) to examine whether participants attempt to find out how the stimulus groups differ; and (2) to investigate the joint influence of impression formation and reinterpretations on the perception of illusory correlation.

With respect to the process of differentiation between the stimulus groups, we specified two stages of the process of active searching for differentiation between groups. The first stage refers to a *general search* involving both the meaning of the task and/or the existence of a relation between the groups and behaviours. We expected that this search would result in an expectation or hypothesis involving the relation between group membership and behaviour. The second stage involves *testing this hypothesis* which might be followed by finding the relation, or 'fit', between group membership and desirability of behaviour.

From the process of intergroup differentiation we derived two predictions involving the perception of illusory correlation. The first prediction is that after the general search, participants would develop the hypothesis that Group A is 'better' than Group B. This hypothesis is based on the comparative fit, caused by the real differences between the groups in the skewed stimulus distribution. Participants were then expected to test this hypothesis (prediction 2). If they were looking for confirming evidence, their hypothesis would be supported, resulting in a definite judgment of a fit, and as a consequence illusory correlation should occur (prediction 2a). On the other hand, we predicted that focusing on disconfirming rather than confirming evidence should have led to the rejection of the hypothesis. This is expressed in terms of a 'non-fit' (i.e. both groups are equally positive) or a 'negative fit' (i.e. Group B is better than Group A), and as a consequence illusory correlation should disappear (prediction 2b). In other words, the perception of illusory correlation should depend on the kind of testing strategy employed such that a confirming strategy would result in illusory correlation effects as opposed to a disconfirming strategy.

Finally, we expected that impression formation and reinterpretation would influence and reinforce each other. As a consequence, we predicted that the search for intergroup differentiation and the resulting illusory correlation (as measured by the standard tasks) would be correlated positively with accentuation effects in the evaluations of the perceived stimuli (prediction 3).

Method

Participants and design

The participants were 22 students of the University of Amsterdam who were paid for their participation. In this study[5] there was one group of participants.

Stimulus materials and procedure

In contrast to Study 1, participants were presented with the whole stimulus series as in the standard illusory correlation task. They were shown 24 statements about Group A members, of which 16 were positive and 8 were negative, and 12 statements about Group B members, of which 8 were positive and 4 were negative. There was no relation between behaviour and group membership, because the ratio of positive to negative statements was identical for both groups (2:1). The procedure differed from Study 1 in certain respects. In this study the participant sat in front of a personal computer and the statements were presented at a rate of 12 s per item. We increased the presentation time simply because verbalization takes time (Ericsson & Simon, 1984). The experiment was conducted by two experimenters, and in each session there was one experimenter and one participant. The experimenter was seated next to the participant and asked him or her to read aloud all information presented on the computer screen and to think aloud during the whole session. This was recorded on audiotape. In order to practise thinking aloud, participants started by solving a few puzzles and were stimulated by the experimenter to say everything that came to mind. When the illusory correlation experiment started, the experimenter only intervened with short standardized prompts (e.g. 'Please, try to say what you think') when a participant did not think aloud during three successive statements.

Another difference from Study 1 was in the nature of the dependent measures. Because we wanted to investigate whether the process of active searching for differentiation also occurs in the standard paradigm, we used the traditional assignment task in which participants again received all the statements they had been presented with before, but without information about group membership. In line with traditional research, we also changed the order of the dependent measures as compared to Study 1. Participants completed the assignment task first, followed by the estimation task and trait ratings. Thereafter, a short unstructured interview followed in which the experimenter could ask the participant for help in clarifying utterances that were not comprehensible for the experimenter. Moreover, the participant was invited to give his or her experiences and impressions of the experiment.

Protocols. The protocols were scored for the search for differentiation as well as other strategies (e.g. distinctiveness of Group B) on each of the four tasks, that is, from the perception of the stimuli to the final task (evaluative ratings). There was also a category for utterances that did not fall into a particular strategy, the so-called 'remaining category'. Thus any utterance was coded, making the system as complete as possible. This was done by the two experimenters, who also conducted the experiment, by coding independently the recorded verbal utterances of each participant. One of these coders was blind to the hypotheses tested in this study. The search for differentiation was divided in two succeeding stages. The coders noted any deviation from this order. The first stage refers to a *general search* involving either the meaning of the task (e.g. 'What am I supposed to do? I think I must . . .'), or the general relation between groups and behaviours (e.g. 'I'm looking for some connection . . .'). This general search was scored dichotomously (either the corresponding behaviour occurred or not) on each task. Next, the judges noted the developed impression or *hypothesis* about group membership and behaviour according to the following seven categories: Group A is more positive; Group B is more negative; Group A is positive and Group B is negative; Group A is more negative; Group B is more positive; Group A is negative and Group B is positive; no hypothesis or other hypothesis. The categories were scored dichotomously (either the corresponding hypothesis occurred or not) on each task.

[5]One might object to this study and the findings because of the small sample size. Because of this small size, we have formulated our conclusions tentatively.

The second stage concerns *hypothesis testing*. This assessed whether participants tested their hypothesis by means of confirmations and disconfirmations. For example, the response 'That is typical Group A behaviour because Group A is . . .' (when perceiving a statement) is judged as a confirmation, and 'I did not expect that from Group B because Group B is . . .' is judged as a disconfirmation of a hypothesis. Both judges assessed whether confirming utterances occurred more often than disconfirming utterances, and they noted the most frequent type of response (i.e. either confirmation or disconfirmation) on each task. We have argued that hypothesis testing might result in attempts to find a *fit* between group membership and desirability of behaviour. Fit was judged according to the following three categories: Group A is more positive; Group B is more negative; Group A is positive and Group B is negative (the first two categories can be considered as an implicit fit, and the third category as an explicit fit). Responses indicating an equal proportion of positive and negative behaviours in both groups were classified as a'non-fit'. Judgments of 'negative fit' were distinguished as follows: Group A is more negative; Group B is more positive; Group A is negative and Group B is positive. Judgments of (non/negative) fit were scored dichotomously on each task.

Reinterpretations. Reinterpretations were measured during the whole experimental session, and computed by the following score index: all utterances involving a negative evaluation of Group A were summed and this was subtracted from all utterances involving a positive evaluation of Group A. Thus a higher score index refers to a more positive interpretation of Group A. The range of the scores varied from −6 (negative Group A interpretation) to +7 (positive Group A interpretation), with 0 indicating 'net neutral reinterpretations'. A negative evaluation of Group A resulted from summing the four following interpretations: both positive and negative behaviours of Group A that were interpreted relatively more negatively, plus both positive and negative behaviours of Group B that were interpreted relatively more positively. A positive evaluation of Group A resulted from summing the four following interpretations: both positive and negative behaviours of Group A that were interpreted relatively more positively, plus both positive and negative behaviours of Group B that were interpreted relatively more negatively. For example, the supposedly negative behaviour of a Group A member that 'comments loudly on the movie in the cinema' was coded as less negative when a participant responded on this statement with 'that is human, and can be very funny'. Similarly, the supposedly positive behaviour of a Group B member 'who organizes a party for the 25th wedding anniversary of her/his parents' was coded as less positive by the utterance (of the same participant), 'that's overdone by that Group B member, the parents could do that for themselves'. Thus, for this participant the reinterpretation score was two, referring to a positive evaluation of Group A.

For each participant a reinterpretation score was computed on each of the four tasks, that is, from the perception of the stimuli to the final task (trait ratings). This was done by the two experimenters who also conducted the experiment, by independently coding the recorded verbal utterances of each participant.

Results

Three participants were excluded from the analyses because they found it embarrassing to think aloud, which resulted in very incomplete protocols.

Agreement between judges

The degree of agreement between the scores of the judges was computed on each task. This was done for all strategies (e.g. search for differentiation) and for the remaining category. The judges agreed that remaining category was only used in the very beginning of the perception stage and that about 20% of all utterances fell into this category. An example of such an utterance is: 'So, I have to read all the statements, that's my task'. Both judges did not report any search for differentiation on the estimation task and evaluative ratings task. This finding supports our

expectation that the search for differentiation occurs in the beginning of the experimental session. Thus we present below only the results for the perception phase and the assignment task. The degree of agreement between the judges varied from 75% to 95% (M = 86%) for all variables in the perception phase. For the variables on the assignment task the agreement varied from 67% to 92% (M = 77%). Thus, the degree of agreement between the judges was reasonably high.

The degree of agreement between the reinterpretation scores of the judges was 84% for the perception stage, and 95% for the assignment task. The judges did not report reinterpretations on the estimation task and trait ratings. Moreover, they agreed that the overall positive evaluation of Group A consisted of negative Group A behaviours that were seen as less negative, *and* of positive Group B behaviours that were seen as less positive. The judges found no evidence for positive Group A behaviours that were interpreted more positively, or for negative Group B behaviours that were interpreted more negatively. For those participants who had an overall negative evaluation of Group A, the judges agreed that this evaluation consisted always of positive Group A behaviours that were seen as less positive, *and* of negative Group B behaviours that were seen as less negative. Thus, there was no evidence for negative Group A behaviours that were interpreted more negatively, or for positive Group B behaviours that were interpreted more positively.

Search for differentiation: stage 1. Four participants tried to find out the *meaning* of the task in the perception phase, for example by saying 'I think I've got to remember which behaviour belongs to which group, but there are too many . . .'. None of the participants reported this search on the assignment task. Six participants reported searching for a *general relation* between the stimulus groups and the behaviours during the perception of the statements, and another five participants indicated a similar search at the beginning of the assignment task. The following reaction in the perception phase illustrates this search: 'I'm sure that there is a sort of system in this . . . but I still haven't found it'. Taken together, 15 participants (79%) were engaged in searching for the meaning of the task or the relation between groups and behaviours.

Developing a hypothesis

The impression that 'Group A is better than Group B' was reported by four participants during the perception of the statements, and by two other participants on the assignment task. A number of participants mentioned one or two impressions that imply a comparison between the groups. In the perception phase, five participants referred to the implicit comparison 'Group A is relatively positive' and another two participants did so on the assignment task. The other implicit comparison 'Group B is relatively negative' was mentioned by one participant in the perception stage, and by two participants on the assignment task. However, with respect to the latter comparison, two participants had first reported that 'Group A is relatively positive'. For the other participant this order was reversed. In sum, 13 participants (68%) generated (either implicitly or explicitly) the hypothesis that 'Group A is better than Group B'. Only one participant mentioned the reversed hypothesis (i.e. Group B is 'better' than Group A).

Active search for differentiation: stage 2. In the perception phase, four participants looked for confirming evidence and three participants reported disconfirming evidence for their hypothesis. On the assignment task no one referred to confirmations, but another seven participants mentioned disconfirmations of their hypothesis. An illustration of the latter is given by a participant who had indicated during the perception phase that 'Group A behaves predominantly well as compared to Group B'. After reading a statement describing undesirable behaviour on the assignment task, the participant asserted: 'Well, I think that Group A members are not always nice, therefore his unkind behaviour must be theirs'. Taken together, 14 participants (74%) tested their hypothesis.

Fit between group membership and behaviour

We distinguished above between an implicit and explicit fit. There were no participants who formulated an implicit fit after testing their hypothesis in the perception phase and on the assignment task. During the perception of statements seven participants reported an explicit fit involving a more positive evaluation of Group A as compared to Group B. Another three participants expressed this fit on the assignment task. Taken together, 10 participants (53%) reported a fit.

Four participants indicated a non-fit between group membership and behaviour by expressing the view that both groups were equally positive. Three participants mentioned a negative fit such that Group B is better than Group A, and two participants gave no response.

Interview. Some relevant and additional information was obtained in the interviews at the end of the experimental session. This was particularly the case for three participants who found a fit without mentioning a preceding search for a relation between the groups and behaviours, or without generating a hypothesis. Two of them reported the fit in the beginning of the perception phase. Explaining this in the interview, one of them said, 'I understood very quickly that one group had to be better and I was convinced after about seven statements that it was Group A'. The other participant stated that: 'After seeing some behaviours, I thought it was about friendly, social behaviours and unfriendly behaviours. Because I saw more friendly behaviours and I also knew that Group A is larger than B, it was quite obvious that Group A must be the good group and Group B the bad one'. Both reactions seem to indicate a preceding search ('I understood very quickly that . . .' and 'I thought it was about . . .'), although it was not articulated during the session. Moreover, it also seems very likely that both participants found (tacitly) confirmations in the remaining statements. The third participant reported a fit at the beginning of the assignment task, also without referring to a search or hypothesis testing. In the interview this participant explained: 'after perceiving about half of the statements, I discovered that Group A was better than B, but I did not want to say this aloud because it felt like stereotyping. Therefore, I waited till all statements were presented, but it did not change my opinion'. The phrase '. . . I discovered' seems to refer to a search, and 'but it did not change my opinion' implies that the statements confirmed the participant's hypothesis.

Table 4. Illusory correlation effects (Study 2)

Assignment task		Estimation task		Trait rating	
Positive A	15.95 (16)	Positive A	16.16 (16)	Group A	25.74
Negative B	7.00 (4)	Negative B	5.89 (4)	Group B	18.74
*Z*phi	.31**	*Z*phi	.18*	Difference	7.00*

*$p < .05$; **$p < .01$.
Note: Level at which mean is different from zero (based on one-tailed *t* tests).

In conclusion, although these participants did not articulate a search for a relation between the groups and behaviours or test their hypothesis, it seems reasonable to conclude that they did search before reporting a fit, and then looked for confirming evidence.

Illusory correlation. For the trait rating task we first recoded the rating scales for friendly–unfriendly and reliable–unreliable. The four trait ratings yielded a Cronbach's alpha of .85 for Group A and .87 for Group B. We computed for each group an evaluative-index score (range from 4 to 36; a higher rating indicates a more positive evaluation). Furthermore, the phi coefficient was converted to a Fisher's *Z* score after the distributions of phi and *Z* were found to be generally normal. Table 4 reports the results on all measures of illusory correlation.

In line with our prediction and as in Study 1, Group A was judged more positively than Group B on all three measures of illusory correlation.

The process of searching for differentiation, reinterpretation and illusory correlation. Our first prediction was that a general search for the task meaning or for a relation between group membership and the valence of behaviours would lead to the hypothesis that Group A is 'better' than Group B (because of the comparative fit). Furthermore, we expected (prediction 2) that subsequent tests of this hypothesis would affect the discovery of a (non- or negative) fit and consequently result in a corresponding illusory correlation. In this process of differentiation, we expected that evaluative reinterpretations would be related to the perceived illusory correlations (prediction 3). In order to examine these predictions, it is necessary to describe the entire process for the individual participants.

Including the three participants mentioned under 'Interview' above, 16 participants (85%) generated (either implicitly or explicitly) the hypothesis that Group A is better than Group B. The findings for these participants involving the search for differentiation and the perceived illusory correlation are presented in Table 5.

First, it is important to note that these participants followed the theoretically based order of a general search, hypothesis testing and finding a fit between group membership and behaviour. The results in the first two rows supported our first prediction that after a general search, the hypothesis 'Group A is better than B' emerged, and that looking for confirming evidence for this hypothesis resulted in a fit and subsequently in corresponding illusory correlations (prediction 2a). At a

Table 5. Process of differentiation, reinterpretation and illusory correlation for participants with the hypothesis 'Group A is better than B' (Study 2)

					Illusory correlation		
Partic. *n*	General search	Hypoth. testing	Reinterpretation[a]	(Non- or negative) Fit	Estimat[b]	Assign[b]	Eval. ratings[c]
3	yes	confirm.	2.67	A + B −	.52 (.34)	.80 (.32)	16.00 (11.79)
3	yes[d]	confirm.	2.67	A + B −	.41 (.16)	.72 (.65)	16.67 (11.37)
4	yes	disconf.	3.75	A + B −	.40 (.22)	.54 (.26)	14.25 (9.07)
4	yes	disconf.	−.25	A & B equal	.00 (.00)	−.02 (.05)	−.25 (.50)
1	yes	disconf.	−6.00	A −B +	−.28	.04	−5.00
1	yes	disconf.	.00	no response	.08	−.04	−2.00

[a]Mean refers to the evaluative reinterpretations during the presentation of stimuli.
[b]Mean *Z*phi scores with standard deviation in parentheses.
[c]Scale is from −32 to +32, a higher positive mean indicates a more positive Group A evaluation. Standard deviation is given in parentheses.
[d]Although not articulated, we have argued (see paragraph 'Interview') that these participants did search for a relation and for confirmations.

first glance, the findings in the third row seem not to support this prediction, because being attentive to disconfirming evidence should have resulted in a non- or negative fit. However, these participants ultimately reported a fit and also corresponding illusory correlations. A plausible explanation for this contradiction is found in the frequency of disconfirmations. Each of the participants in the third row mentioned disconfirming evidence for their hypothesis only twice, whereas other participants (in the last three rows) mentioned this more often (four–eight times). Thus, it seems that the low frequency of disconfirmations was not sufficiently convincing to reject the hypothesis which is subsequently expressed in the fit, and consequently produced positive illusory correlation.

Support for prediction 2b is found in the last three rows. We predicted that testing the hypothesis 'Group A is better than B' by focusing on disconfirming evidence should result in a non- or negative fit, and as a consequence illusory correlations should disappear. A non-fit was expressed as 'both groups are equally positive' in the fourth row, and no illusory correlation occurred. The participant in the fifth row reported a negative fit in terms of 'Group B is better than Group A', and displayed corresponding negative illusory correlation on two tasks. Although the participant in the final row did not articulate any fit, it seems that the perception of disconfirming evidence resulted in the rejection of the hypothesis by showing hardly any illusory correlations.

There were three participants whose responses did not support our predictions. One participant neither searched for a general meaning nor tested a hypothesis. On the assignment task, this participant said: 'I think I must decide which group is good and which one is bad, but I really don't know. I guess A is bad and B is good, but my decision is completely arbitrary'. Another participant searched and generated the hypothesis that Group B is better than A. After looking for confirming evidence a corresponding fit was mentioned which should have resulted in negative illusory correlation scores (see also the fifth row in Table 5). However, the responses on the illusory correlation tasks revealed no illusory correlations. The third participant did not articulate any sign of a search for differentiation and did not display illusory correlation.

All in all, our predictions concerning the process of searching for intergroup differentiation and the resulting illusory correlations are supported by 16 (84%) participants.

Reinterpretations. Participants' utterances involving the evaluative reinterpretations of behaviours performed by the groups were compared to pretest ratings. In line with Berndsen *et al.* (1996, 1998), we investigated whether the reinterpretations favouring Group A began during the presentation of the stimuli. Comparing the reinterpretation score to zero (indicating no changes) revealed that these reinterpretations occurred in the perception stage ($M = 1.42$, $t(18) = 2.23$, $p < .05$), but not in the assignment task ($M = .16$, $t(18) = .82$, n.s.). It is also interesting to note that 16 of the 19 participants (84%) reinterpreted statements during the perception stage (the number of reinterpretations varied from one to eight). Three participants started to reinterpret after perceiving the fourth statement, seven participants started between the ninth and fifteenth statement, and six participants

after the nineteenth statement. In other words, evaluative reinterpretations are more likely to occur during the perception of the stimuli.

As in the research of Berndsen *et al.* (1996, 1998), we divided the presented statements into two halves. Both sets consist of eight A +, four A −, four B + and two B − items (thus there was no correlation between group and behaviour), and we compared the reinterpretation scores in both sets. The degree of reinterpretations differed marginally significantly between the two sets of items ($t(18) = 1.69$, $p < .06$, one-tailed), indicating that positive Group A and negative Group B interpretations tended to increase over the course of perceiving the statements (first half: $M = .37$; second half: $M = 1.05$).

With respect to the entire process of differentiation and reinterpretation, the first three rows of Table 5 demonstrate that the search for intergroup differentiation corresponds to the nature of the evaluative reinterpretations during the presentation of the stimuli (col. 5) and the resulting illusory correlations. The number of positive reinterpretations indicates a positive Group A interpretation which is in line with the subsequent positive illusory correlation effects. The actions of participants who focused on disconfirmations of the hypothesis that Group A is better than Group B and reported a non-fit (row 4) or a negative fit (row 5) are also in accordance with the respectively negligible number of reinterpretations and negative reinterpretations, which are in line with the subsequent illusory correlations. These findings support our third prediction involving the relation between the search for intergroup differentiation, evaluative accentuations of the stimuli and the perception of illusory correlation.

We expected that forming an illusory correlation and reinterpreting the evaluations of the statements would be not independent, but reciprocal and reinforcing processes. The following representative reactions can illustrate this: perceiving desirable behaviour from a Group A member, a participant asserted: 'that's typical for Group A; they are nice and always do the right things'. Reading the (supposedly negative) statement 'A member of Group A often doesn't finish things he/she just started work on', the same participant replied: 'That must be a very dynamic person, the people belonging to Group A are spontaneous and interested in so many things'. Another illustration of the mutual reinforcement of forming impressions and reinterpreting is the following example. A participant started to form a negative impression of Group B. Perceiving the statement 'A member of Group B is polite to his/her guests', the participant evaluated this as 'that person will be over-polite, and that's annoying'. A few statements later a statement appeared that described undesirable Group B behaviour, and the participant said: 'What else could I expect from Group B? They are unkind and anti-social'.

Moreover, the finding that the illusory correlation effects were related to accentuation effects in the evaluations of the perceived stimuli supported our third prediction.[6] The correlations between the scores on the three illusory correlation measures and the reinterpretation score during both the perception of stimuli and

[6]There was one outlier because the reported reinterpretations involved a positive evaluation of Group A in contrast to the illusory correlation effects that favoured Group B. This participant became confused in the assignment task and said that the choice to favour Group B as opposed to Group A was arbitrary. We excluded the participant from the analysis.

the assignment task were significant (reinterpretations in the perception stage assignments: $r = .50$. $p < .05$; frequency estimates: $r = .73$, $p < .01$; trait ratings: $r = .61$, $p < .01$; during the assignment task assignments: $r = .54$, $p < .05$; frequency estimates: $r = .51$, $p < .05$; trait ratings: $r = .51$, $p < .05$).

Discussion

The second study demonstrates that the evaluative reinterpretations and the search for intergroup differentiation start during the presentation of the statements in the standard paradigm. Moreover, it shows that the occurrence of illusory correlation depends on the search for differentiation, and particularly on the strategy used to test the differences between the two groups. Looking for confirming evidence leads to the perception of illusory correlation, while focusing on disconfirming evidence produced no (or negative) illusory correlations. In line with Study 1, participants also differentiated between the stimulus groups by interpreting both the positive behaviours performed by Group B as relatively negative, and the negative Group A behaviours as relatively positive. Moreover, this differentiation tendency is positively related to the size of the illusory correlation effect. The correlations between the illusory correlation measures and the reinterpretations replicate both the findings of the first study and those of Berndsen *et al.* (1996, 1998). Our second study also provides an answer to the question of how these two kinds of differentiation—the search for intergroup differentiation and the evaluative reinterpretations—are related. Findings show that the reinterpretations serve to reinforce the hypothesis that Group A is better than Group B, and that they are complementary to a confirmatory hypothesis testing strategy.

Finally, the second study provides more insight into the *relation between illusory correlation and the interpretive process* than the first study. Because the observed reinterpretations in the first study were measured *after* the illusory correlation tasks, they can be seen as another measure of illusory correlation, even though Berndsen *et al.* (1996, 1998) showed that reinterpretations (measured by the rating task) occurred *before* completing the illusory correlation tasks. The point to make here is that the second study supports the idea that evaluative reinterpretations and the perception of illusory correlation influence and reinforce each other. First of all, we observed relationships between the illusory correlation scores and reinterpretations during the perception of the stimuli. Next, we also showed that the majority of the participants reinterpreted the evaluations of the behaviours during the initial presentation stage, and that half of them started to do that after having perceived about one third of the statements. Further support involves participants' verbal reports indicating that when they have developed the hypothesis that Group A is better than Group B, they subsequently reinterpreted the disconfirming behaviours in accordance with their hypothesis and generally commented upon confirming behaviours as providing support for their hypothesis. This then should have reinforced the impression that Group A is better than Group B. More importantly, the latter idea is confirmed by the increase in overall positive Group A interpretations in the second half of the perception stage as compared to the first

half. In short, the joint influence of impression formation and evaluative reinterpretations occurred when statements are perceived, leading to a strengthening of the original impression which resulted in subsequent illusory correlation effects.

Although it is generally known that verbalizing thoughts can influence the thinking process *per se,* the observed replications of previous results seem to justify the conclusion that the thinking-aloud method did not interfere materially with the perception of illusory correlation as measured in traditional research.

Both studies extend the work of Berndsen *et al.* (1996, 1998) by demonstrating *which* kinds of stimuli are subjected to reinterpretations. Overall, the disconfirming rather than the confirming behaviours are reinterpreted (e.g. given a positive Group A impression, undesirable Group A and desirable Group B behaviours were reinterpreted).

The observation that participants reinterpreted only the disconfirming rather than the confirming behaviours into supporting evidence for their impression suggests that justifications may also play a role in imposing restrictions on the scope of reinterpretations. This is in accordance with Kunda's (1990) work which shows that drawing a particular conclusion is affected by justifications for this conclusion. Our findings could also be related to research on the 'confirmation bias', that is, the tendency to confirm rather than disconfirm hypotheses by maximizing the probability of a confirming outcome (Klayman & Ha, 1987). The second study supports the confirmation tendency, and also reveals how participants search for confirmation. That is, they do not focus on statements that confirm their impression, but rather on reinterpreting disconfirming behaviours, resulting in support of their impression.

Finally, it is interesting to note that there is not much evidence for the salience of infrequent information as proposed by the distinctiveness account. In that case one should expect that participants generate a hypothesis in which negative Group B behaviour is mentioned explicitly. However, only three participants did this, while two of them referred earlier to a hypothesis involving the positivity of Group A.

General discussion

The contribution of the present research is that it shows the role of interpretive processes underlying the illusory correlation effect in the presence and absence of distinctive stimulus information.

We argue that imposing sense on the stimulus situation can lead to illusory correlation. This kind of meaning enhancement can be considered as understandable and rational behaviour in the context of the illusory correlation task. However, one could argue that even though the process leading to illusory correlation is rational, the product itself is biased. This argument is in accordance with the widespread opinion that the absence of illusory correlation reflects adequate information processing. However, our second study offers a different explanation: Participants who reported no illusory correlations started with the

same hypothesis (i.e. the positivity of Group A over Group B) as participants who displayed illusory correlations. The only difference between them was the test strategy; the former (smaller) group of participants focused on disconfirmations (resulting in attenuated illusory correlations), whereas the latter searched for confirmations. In other words, judgments of the absence of a correlation apparently do not result from 'accurate' perception, but from a different test strategy. The point to make here is that either the perception of illusory correlation or its absence can arise from the same processes involving the search for differences between groups, suggesting that a sense-making process of differentiation plays a central role in the illusory correlation paradigm. In focusing on sense-making processes, the differentiated meaning account is perhaps the only major explanation of illusory correlation that does not rely on (biased) memory processes. Recent evidence from a source monitoring analysis supports the notion that illusory correlation is not explicable in terms of memory processes but reflects "response bias' (Klauer & Meiser, 2000). The present approach is consistent with this finding.

With respect to the practice of conducting research, we believe that sense-making processes are not necessarily confined to the illusory correlation task. We expect that such processes occur whenever experimental settings create expectations (other than originally conceived by the researchers) which affect the psychological reality of the task and consequently the responses.

The process of deriving and giving meaning fits in very well with self-categorization theory (Turner *et al.*, 1987) which views categorization as an adaptive process of making sense of the world, and shows that participants can make sense of situations by actively reinterpreting their evaluations of the stimuli so that they become categorically meaningful. How does this sense-making process relate to the process of stereotyping about social groups? A prevailing view of stereotypic beliefs is that they result from cognitive information processing biases. This is based on the assumption that the information in our environment is too complex for our processing capabilities, hence the need to simplify this information, resulting in stereotypes (see e.g. Hamilton, 1981). According to this view, stereotypes are erroneous generalizations based on distorted or illusory perceptions of individuals as group members. However, the present studies, if anything, suggest the opposite, namely that illusory correlations result from cognitive effort to understand and enhance the meaning of the task. Extending this idea to the area of stereotypes, this implies that stereotypes are not so much the product of simplification but of effortful elaboration.

Acknowledgements

This project was funded by the Dutch Organization for Scientific Research (NWO), SGW grant nr. 575-70-054. We would like to thank Kim van Baaren for her assistance in conducting Study 2 and coding the audiotapes.

References

Abele, A. E., & Petzold, P. (1998). Pragmatic use of categorical information in impression formation. *Journal of Personality and Social Psychology 75*, 347–358.

Acorn, D. A., Hamilton, D. L., & Sherman, S. J. (1988). Generalization of biased perceptions of groups based on illusory correlations. *Social Cognition, 6*, 345–372.

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.

Berndsen, M., Spears R., Van der Pligt, J., & McGarty, C. (1999). Determinants of intergroup differentiation on the illusory correlation task. *British Journal of Psychology 90*, 201–220.

Berndsen, M., Van der Pligt, J., Spears, R., & McGarty, C. (1996). Expectation-based and data-based illusory correlation: The effects of confirming versus disconfirming evidence. *European Journal of Social Psychology, 26*, 899–913.

Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversations. *European Journal of Social Psychology, 23*, 149–165.

Carroll, L. (1960). *Through the looking-glass.* Norwich: Fletcher, & Son.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis. Verbal reports as data.* Cambridge, MA: MIT Press.

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., Hemmeter, U., & Hofmann, C. (1984). On the origin of illusory correlations. *European Journal of Social Psychology, 14*, 191–201.

Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory correlations and memory performance. *Journal of Experimental Social Psychology, 29*, 111–136.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 83–115).

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

Hamilton, D. L. (Ed.) (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.

Haslam, S. A., & McGarty, C. (1994). Problems with the measurement of illusory correlation. *European Journal of Social Psychology, 24*, 611–621.

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.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118*, 248–271.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Klauer, K. C., & Meiser, T. (2000). A source monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin, 26*, 1074–1093.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

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., & de la Haye, A. (1997). Stereotype formation: Beyond illusory correlation. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 144–170). Oxford: Blackwell.

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.

McGarty, C., & Turner, J. C. (1992). The effects of categorization on social judgement. *British Journal of Social Psychology, 31,* 253–268.

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology, 29*, 11–28.

Oakes, P. J., Turner, J. C., & Haslam, S. A. (1991). Perceiving people as group members: The role of fit in the salience of social categorizations. *British Journal of Social Psychology, 30*, 125–144.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.
Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion–exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 217–245). Hillsdale, NJ: Erlbaum.
Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology, 27*, 107–123.
Stapel, D. A., Reicher, S. D., & Spears, R. (1995). Contextual determinants of strategic choice: Some moderators of the availability bias. *European Journal of Social Psychology, 25*, 141–158.
Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues, 25*, 79–97.
Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford: Blackwell.

Received 28 April 1999; revised version received 13 December 1999

Appendix

On the frequency estimation task the expected values in the standard condition differ from the other conditions, because in the latter conditions Groups A and B were supposed to be larger than in the standard condition in order to keep the total number of presented stimuli constant for all conditions. With regard to the assignment task, all participants were presented with 12 positive behaviours and six negative behaviours. Given participants' knowledge that Group B is smaller than Group A, we computed the possible expected values involving *no correlation* between group membership and behaviour. That is, with a total number of 18 statements the size of Group B must be smaller than nine and that of Group A must be greater than nine. The table below reports the three possible expected values distributions:

Three distributions of expected values in the assignment task

	Positive	Negative	Positive	Negative	Positive	Negative
Group A	8	4	10	5	12	6
Group B	4	2	2	1	0	0
	12	6	12	6	12	6

It can be seen that the expected values for positive behaviour in Group A could be 8, 10 or 12 and that the corresponding expected values for negative behaviour in Group B could range from 0 to 2.