Expectation-based and data-based illusory correlation: the effects of confirming versus disconfirming evidence

Berndsen, M.; van der Pligt, J.; Spears, R.; McGarty, C.

Published in:
European Journal of Social Psychology

DOI:
10.1002/(SICI)1099-0992(199611)26:6<899::AID-EJSP795>3.0.CO;2-B

UvA Keur
UB Groningen
Broerstraat 4
9700 AN Groningen

A078294789 ISN: 681027 RAPDOC (R) **OPGEHAALD** NCC/IBL

Verzoeke te behandelen voor: 25-04-2005 **Ingediend door:** 0004/9999
Datum en tijd van indienen: 11-04-2005 18:47 **Datum plaatsen:** 11-04-2005 18:47 **Type aanvrager:** UKB **I.D.:** UVA KEUR (UB GRONINGEN)

PPN: 790215462

European journal of social psychology 1971 Den Haag [etc.] Mouton
Gewenst: 1996-00-00 **Deel:** 26 **Nummer:** 6 **Electronisch leveren (LH=N)**

Auteur:	**Titel van artikel:**	**Pagina's:**
Berndsen, M. (ed.)	Expectation-based and Data-based Illusory Correlation: The Effects of C	899-914

Opmerking:
arno ID: 116480

LEEUW	P0312	Vol. (1971)-19(1989),21(1991)-32(2002)
WWW		vol. 27(1997)-

1. ○ origineel gestuurd
2. ○ fotokopie gestuurd
3. ○ overige
4. ○ nog niet aanwezig
5. ○ niet aanwezig
6. ○ niet beschikbaar
7. ○ uitgeleend
8. ○ wordt niet uitgeleend
9. ○ bibliografisch onjuist
10. ○ bij de binder

Fakturen zenden aan: Rijksuniversiteit Groningen
Bibliotheek, Uitleenbureau
Postbus 559
9700AN Groningen

European Journal of Social Psychology, Vol. 26, 899–913 (1996)

Expectation-based and data-based illusory correlation: the effects of confirming versus disconfirming evidence

MARIËTTE BERNDSEN, JOOP VAN DER PLIGT and RUSSELL SPEARS
Department of Social Psychology, University of Amsterdam, The Netherlands

and

CRAIG MCGARTY
The Australian National University, Australia

Abstract

The present study ($n = 154$) examines the effects of expectations and stimulus information on the perception of illusory correlation. There have been few studies attempting to integrate expectation-based and data- (distinctiveness-) based processes. These studies suggest that data-based illusory correlation can be overruled by prior expectations, but it is not clear whether this is a consequence of a confirmation bias. In the present study, where participants were not exposed to the specific stimulus information, expectation was manipulated by stating that group B behaved more negatively than group A. Moreover, participants were provided with information contained in a statement-rating task that allowed for the confirmation and disconfirmation of the prior expectations. Participants rated the desirability of these behaviours and also performed the standard illusory correlation tasks. Based on self-categorization theory and Alloy and Tabachnik (1984), we predicted that in the absence of prior expectations, completing the rating task before the illusory correlation tasks would produce stronger illusory correlation than the reverse order. However, in the presence of prior expectations we expected the rating task to undermine illusory correlation, because the information obtained in this task tends to disconfirm prior expectations. Results support the predicted interaction between task order and expectation. We discuss some implications for research on confirmation bias.

Addressee for correspondence: Mariëtte Berndsen, Department of Social Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands, e-mail: sp_berndsen@macmail.psy.uva.nl.

This project was funded by the Dutch organization for scientific research (NWO), SGW grant nr. 575-70-054.

Received 21March 1995
Accepted 25 December 1995

CCC 0046-2772/96/060899–15

INTRODUCTION

Research on the illusory correlation phenomenon, the tendency to perceive a covariation between two classes of stimuli that are uncorrelated or less strongly correlated than perceived, has been widely researched in social psychology. There has been research on both data-based (what is often termed *distinctiveness-based* illusory correlation) and expectation-based illusory correlations. Unfortunately, there have been few attempts to integrate these two lines of research. The present study deals with the *relative* influence of expectations and empirical data on the perception of illusory correlation. First, we will briefly describe the evidence concerning data-based illusory correlation, followed by evidence concerning theory-based, or expectation-based illusory correlation.

The term *illusory correlation* was originally coined by Chapman (1967) to describe the overassociation of semantically related word pairs or word pairs of unusual length. Hamilton and Gifford (1976) applied the concept of illusory correlation to the perception of social groups and developed a now familiar paradigm to demonstrate this effect. They presented participants with behavioural instances about members of two groups, called group A and group B, and group A represented the majority and group B the minority. These two groups exhibited the same ratio of desirable to undesirable behaviours (9:4, Study 1). Thus there was no correlation between type of behaviour and group. Hamilton and Gifford demonstrated that the co-occurrence of the infrequent (negative) behaviour and the infrequent group (group B) was overestimated. They proposed that this data-based illusory correlation occurs because statistically infrequent combinations are particularly distinctive for the perceiver and receive more attention, are more efficiently encoded, and consequently, are more available in memory than non-distinctive categories. Even though this explanation has been challenged by a number of researchers (Fiedler, 1991; McGarty, Haslam, Turner & Oakes, 1993; Smith, 1991), and because distinctiveness is not a factor in these explanations we prefer the term 'data-based', there is substantive support for the data-based illusory correlation effect (see e.g. McConnell, Sherman & Hamilton, 1994b; Mullen & Johnson, 1990). This type of illusory correlation is based on *data* and may be implicated in the development of stereotypes about minorities as opposed to illusory correlation that is based on *expectations or theories*. The latter may offer an explanation for the maintenance of stereotypes.

Expectation-based illusory, correlation has been investigated for socially meaningful groups (see e.g., Hamilton & Rose, 1980; Spears, Eiser & Van der Pligt, 1987). In these studies there was no co-occurrence of infrequent information (as is the case in data-based research) and the observed illusory correlation effects corresponded with the prior expectations held by the participants.

On a more general level some research in social psychology focuses on the role of expectations in social judgment processes, while others pay more attention to data-based processes. For example, Nisbett and Ross (1980) and Crocker (1981) argued that people rely heavily on their expectations in assessing covariations, and that this might result in distorted judgments when these expectations are inconsistent with the data. Expectation-based illusory correlation may be explained by the *confirmation bias*; that is, the tendency to confirm *a priori* expectations and hypotheses by searching for confirming evidence rather than disconfirming evidence (see e.g., Wason, 1960; Klayman & Ha, 1987). Within social psychology, this point of view is clearly stated by

Crocker (1982) and Snyder and Swann (1978). In contrast, Alloy and Tabachnik (1984) proposed that the perception of covariation is a function of the interaction between the strength of prior expectations and the strength of empirical information. Consistent with the ideas of Alloy and Tabachnik is the argument of McGarty *et al.* (1993) that data-based illusory correlation is created by both expectations and data-based properties. The latter is based on the analysis of comparative fit (i.e. the match between the categories and the comparative properties of the stimuli; Oakes, Turner & Haslam, 1991). McGarty *et al.* argued that participants in the illusory correlation paradigm *expect* to find meaningful differences between the stimulus groups and that the *perceived fit* between the size of the stimulus groups and the relative proportion of the positive and negative behaviours, produces the illusory correlation effects.

A critical feature of McGarty *et al.*'s approach is the argument that the perceived covariations are not necessarily erroneous but are in fact potentially valid interpretations of the available data. That is, there is more than one plausible interpretation of the data and expectations help determine which interpretation is chosen. Unfortunately, McGarty *et al.* did not include a manipulation of expectations in their experiments. Hypothesis testing approaches such as that of McGarty *et al.*, imply that people can interpret empirical data such that they reconcile them to existing expectations, but also that they can change their expectations in line with the perceived evidence. Similarly, when considering their own illusory correlation findings McConnell *et al.* (1994b) state: '. . . old information is reconsidered and can change its meaning in the light of new information. . . . Our findings . . . suggest a dynamic organization of interpretive processes, where past experiences are accessed, reconsidered, and reprocessed in the light of new information and experiences' (p. 427).

There have been very few studies that bear on the relation between expectation and data-based effects. With respect to the latter, it seems fruitful to address the *distinctiveness* (statistical infrequency) of particular data, *and* the treatment of these data as a *test* of the validity of prior expectations. The latter is related to the confirmation bias. McArthur and Friedman (1980) examined illusory correlation when prior expectations and shared infrequency were both present. Their participants were white students and the prior expectations involved associative links between demographic groups (whites versus blacks) and the desirability of behaviours. The results showed that data-based illusory correlation only occurred when infrequency and associative links acted in the same direction (for example, when the infrequent undesirable behaviours were paired with the infrequent group of black persons). However, when infrequency and associative links acted in opposition (for example, white persons and the undesirable behaviours were simultaneously infrequent), the data-based illusory correlation was overruled by the associative link between desirable behaviours and white group membership. These findings seem to support Nisbett and Ross's (1980) and Crocker's (1981) view that covariation judgments are based on prior expectations when these compete with the data. However, given that all participants were white students, we would argue that the observed *confirmation bias* effect might be confounded with an ingroup bias effect (i.e., that the participants favoured their own facial group) rather than prior expectations (*cf.* Schaller & Maass, 1989).

An ingroup bias explanation was ruled out in the research of Fiedler, Hemmeter and Hofmann (1984, p. 197). They manipulated distinctiveness in the context of prior

expectations. One condition presented a skewed stimulus distribution, such that infrequency and expectations acted in the same direction. The other condition presented participants with an equal stimulus distribution. Illusory correlation occurred in both conditions, but was not stronger in the skewed distribution condition. Thus it seems that *distinctiveness* received no additional weight in the presence of prior expectations. Moreover, Fiedler *et al.* demonstrated that prior expectations affected the processing of empirical information. In their experiment they used a rating task and asked participants to rate each attitude statement ('liberal' versus 'authoritarian', produced by students or clerks) on an authoritarian–liberal dimension during stimulus presentation. The results showed that the statements were interpreted in line with prior expectations, indicating that the illusion is already effective during the perception (and evaluation) of the stimuli. This can be considered as a tendency to confirm prior expectations, that is, a *confirmation bias*. Furthermore, Berndsen (1994), who also used a rating task after completion of the standard illusory correlation measures, demonstrated a similar confirmation bias. It appeared that participants who perceived illusory correlations between group membership and desirability of behaviours, modified the meaning of the behavioural statements in the rating task in accordance with their illusory correlations. In other words, in the presence of prior expectations (or perceived illusory correlations) the rating task provides a measure of meaning shifts in the statements.

Finally, Haslam, McGarty and Brown (1996) conducted an experiment where the majority and minority groups were represented by right- and left-handed group members (experimental condition), or by members of group A amd B (control condition). Furthermore, the minority of the behaviours was undesirable. Based on the idea of McGarty *et al.* (1993) that illusory correlation is created by expectations of meaningful differences between the stimulus groups, they predicted no illusory correlation in the experimental condition (as opposed to the control condition), because there was no reason for participants to expect that there would be differences between groups. This prediction was supported. In other words, eliminating expectations of differences between social categories, eliminates the *data-based* illusory correlation.

Thus first of all it is important to note that very few studies have simultaneously addressed expectation and data-based processes. These studies indicate that data-based illusory correlation can be overruled by expectations under some circumstances, but it is not clear whether this is only due to confirmation bias effects. In the study of McArthur and Friedman (1980) the confirmation bias effect might be confounded with an ingroup bias effect, while the studies of Fiedler *et al.*, (1984) and Berndsen (1994) seemed to support confirmation bias. Although their results do not necessarily bear on the issue of confirmation bias, the analysis of McGarty *et al.* (1993), Haslam *et al.* (1996) as well as that of Alloy and Tabachnik (1984), is inconsistent with the proposed role of confirmation bias.

Given this state of affairs we need first of all to experimentally create expectations and then provide a mechanism to confirm or disconfirm them through the presentation of data. We induced an *a priori* expectation by informing half of the participants that group B behaves more negatively than group A. The other half of the participants had no prior expectations about the stimulus groups. The mechanism to confirm or disconfirm this expectation was provided by the rating task used by Fiedler *et al.* (1984) and Berndsen (1994), but following McGarty *et al.*'s (1993) procedure we did not

provide participants with stimulus information from which they could first encode an illusory correlation.

With respect to the rating task, participants in our study were presented with behavioural statements ascribed to group A and B (as in the traditional paradigm) and were asked to rate these statements on a desirability scale. In the distribution of the statements in this task there is no correlation between type of behaviour and group membership, and the negative behaviours of group B are distinctive in terms of their statistical infrequency (in other words, a similar distribution to the standard paradigm). The statements in the rating task reflect empirical information and participants who expect differences between the groups can test this expectation against the data in the rating task. In line with Alloy and Tabachnik (1984) and McGarty *et al.* (1993) we expect that the disconfirming evidence (i.e., negative group A behaviours and positive group B behaviours) in the rating task would attenuate the perception of illusory correlation. In order to investigate this we counterbalanced the order of the rating task and the standard illusory correlation measures.

A different use of the statements in the rating task might take place when there are *no* prior expectations. Following McGarty *et al.*, participants in our study did not receive the standard stimulus information as in the traditional paradigm, but we expected that the rating task would enable participants to perceive differences between the groups, leading to illusory correlation effects.

To summarize then, the aim of the present study is to examine the influence of expectations and empirical evidence on the perception of illusory correlation. The manipulations of task order and of expectation lead to the following hypotheses: when participants *expect* negative behaviour group B, and have the opportunity to test this expectation in the rating task, then the illusory correlation effect should be reduced as compared to the reverse task order. When participants have *no expectations*, doing the rating task first allows them to perceive a fit between group size and the desirability of behaviour. This should lead to illusory correlation as opposed to the reverse task order where there is no basis to differentiate between the groups. In other words, we predict an interaction between expectation and task order.

Responses on the rating task are of interest in themselves. First, a possible benefit of this task is that it allows us to investigate the relative weight of data-based illusory correlation in the context of a prior expectation by examining the response patterns in the first and the second half of the rating task. Second, Fiedler *et al.* (1984) showed that the rating task will give rise to the perception of illusory correlation given that participants have prior expectations. However, we expect that in the conditions *without* prior expectations, the scores on the rating task will reveal illusory correlation effects, independent of task order, because the rating enables participants to perceive the fit between group size and the desirability of behaviour.

METHOD

Participants

A total of 154 first year psychology students of the University of Amsterdam participated in the experiment. Participants were randomly assigned to the experimental conditions.

Design and independent variables

The design of the experiment was a 2 (expectation versus no expectation) × 2 (task order) factorial design. *Expectation* was manipulated by informing participants that group B would display relatively more negative behaviour than group A, versus participants not providing this expectation. *Task order* was manipulated by presenting first the three illusory correlation measures and subsequently the rating task, or vice versa.

Stimulus materials and procedure

In a pilot study 93 behaviour descriptions were rated by 25 participants on a 9-point scale ranging from 'very desirable' (1) to 'very desirable' (9). A selection was used in the experiment: 12 moderately positive items and six moderately negative items for the assignment task and 24 moderately positive items and 12 moderately negative items for the rating task.

Using similar instructions to those of Hamilton and Gifford (1976) participants were informed in a questionnaire about the purpose of the experiment. All participants were told that the experiment is about behaviour of people who belong to one of the two groups, labelled A and B, and that in the real-world group B is smaller than group A. As in the 'no-information condition' of McGarty *et al.* (1993), participants were not provided with the standard stimulus information as in the traditional paradigm. Participants in the no-expectation conditions started either the rating task followed by completing the illusory correlation measures (no-expectation/rate-ic condition), or this order was reversed (no-expectation/ic-rate condition). Other participants were provided with the expectation that group B would display relatively more negative behaviour than group A. In one expectation-based condition participants completed first the rating task and thereafter the illusory correlation measures (expectation/rate-ic condition), and in the other condition this order was reversed (expectation/ic-rate condition).

Dependent measures

All participants had to complete four dependent measures.

Rating task

Participants were asked to evaluate 36 behavioural statements ascribed to group A or group B on a 9-point scale ranging from 'very undesirable' (1) to 'very desirable' (9). There was no correlation between type of behaviour and group: 16 members from group A and eight from group B, performed positive behaviours, eight members from group A and four from group B, performed negative behaviours. The scores on the rating task were combined for each participant according the following formula: rating index $=(MA^{+}-MB^{+})+(MA^{-}-MB^{-})$. A positive rating index indicates a more positive evaluation of group A.

Evaluation task

Two positive ('pleasant' and 'sympathetic') and two negative ('unfriendly' and 'selfish') traits were selected. Participants were asked to evaluate group A and B on 9-point scales ranging from 'not at all' (1) to 'extremely' (9). After recoding the trait rating scales for 'unfriendly' and 'selfish', one evaluative index-score was computed for each group. The possible range was 1 to 9; a higher rating indicates a more positive evaluation.

Frequency estimation task

Participants were informed that 24 behaviours were performed by members of group A and 12 by members of group B. They were asked to estimate how many of the statements about members of both groups described desirable and undesirable behaviours. A phi coefficient was computed from each participant's 2×2 contingency table, representing the number of positive and negative behaviours attributed to each group.

Assignment task

This task differed from the traditional assignment task. In the standard illusory correlation paradigm participants again receive all the statements they had been presented with before, but without information about group membership. They are then asked to indicate group membership of the person who performed each of the behaviours. In the present experiment such a task could not be used because participants were not presented with any statements (although they rated them in two of the four conditions before they completed the illusory correlation measures). We therefore developed a 'projective' assignment task in which all participants were exposed to 18 statements of which 12 described positive behaviours and six described negative behaviours (see also Berndsen, 1994). Participants were instructed to guess from which group each statement emanated. As for the frequency estimation task, a phi coefficient was computed from each participant's 2×2 contingency table. In traditional research the phi coefficients on the assignment task and the frequency estimation task are converted to a Fisher's Z score. Following Haslam and McGarty (1994) distributions of phi and Z were inspected and found to be generally normal distributions. However, the assumption of homogeneity of variance was violated for the transformed phi scores on the estimation task but not for the raw phi scores. Therefore, our analyses are based on the phi coefficients.

RESULTS

Illusory correlation measures

The four trait rating scales yielded a Cronbach's alpha of 0.86 (group A) and 0.88 (group B) and one evaluative index-score was computed for each group. Table 1 reports the mean ratings of group A and B for all conditions. It can be seen that in nearly all conditions group A was evaluated significantly more positively than group

B on the evaluation task. The results on the frequency estimation task and assignment task are also presented in Table 1. In each condition the mean phi scores were compared to zero and this score differed significantly from zero on both tasks in all conditions except the no-expectation/ic-rate condition (see Table 1).

The predicted interaction between expectation and task order was significant for all three illusory correlation measures (evaluation task: $F(1,151)=24.36$, $p<0.001$, estimation task: $F(1,147)=35.49$, $p<0.001$, assignment task: $F(1,150)=16.35$, $p<0.001$)[1]. Planned comparisons supported our hypothesis that participants without expectations who rated the statements first, displayed illusory correlation effects as opposed to participants with the reverse task order: evaluation task ($t(68)=2.31$, $p<0.05$, one-tailed), assignment task ($t(68)=1.58$, $p<0.06$, one-tailed), estimation task ($t(65)=1.89$, $p<0.05$, one-tailed). Furthermore, and as predicted, this effect was reversed for participants who expected negative behaviour in group B; they displayed significantly less illusory correlations when they completed the rating task first as compared to participants who completed the illusory correlation measures first: $t(83)=4.69$, $p<0.001$, one-tailed (evaluation task), $t(82)=6.51$,

Table 1. Results on three measures of illusory correlation and ratings of statements for all conditions

	Provided expectation			
	No expectation		Expectation	
	Task order			
	ic–rate $n=36$[a]	rate–ic $n=34$	ic–rate $n=44$	rate–ic $n=40$
Evaluation task				
Mean gp.A	5.31	6.57	7.00	6.66
Mean gp.B	5.36	5.76	3.77	5.35
Difference	−0.05	0.81**	3.23***	1.31***
Estimation task				
Positive A	14.97	16.68	18.14	16.60
Negative B	4.58	4.71	8.07	4.85
phi	0.01	0.09**	0.42***	0.10**
Assignment task				
Positive A	6.33	7.15	10.14	7.90
Negative B	2.47	3.12	4.45	3.55
phi	−0.05	0.11*	0.58***	0.25***
Item rating task				
Index[b]	0.45***	0.52**	−0.17	0.16
First half	−0.28	−0.36	−1.16	−0.81
Second half	1.16	1.41	0.81	1.12

Note. Level at which mean is different from zero (based on one-tailed t tests).
*$p<0.05$; **$p<0.01$; ***$p<0.001$.
[a]Three participants in this condition did not complete the estimation task.
[b]The rating index is compared to zero because in the pilot study this index was equal to zero.

[1]We did not present the main effects of expectation and task order on the measures of illusory correlation because they are not relevant to the hypotheses we were formally testing in our experiment. (For interest they are included here. Expectation: $F(1,147)=39.55$, $p<0.001$ (estimation task), $F(1,150)=38.20$, $p<0.001$ (assignment task), $F(1,151)=44.87$, $p<0.001$ (evaluation task). Task order: $F(1,147)=12.23$, $p<0.01$ (estimation task), $F(1,150)=1.97$, n.s. (assignment task), $F(1,151)=3.51$, $p<0.07$ (evaluation task)).

$p<0.001$, one-tailed (estimation task), $t(82)=4.60$, $p<0.001$, one-tailed (assignment task).

Rating task

We also predicted that the rating task *per se* (independent of task order) would show perceptions of illusory correlation in the conditions *without prior expectations*. This prediction was supported; the significant overall rating indices in Table 1 indicate that participants differentiated between the groups such that the large group (A) was associated with positive behaviours. In order to get more insight into the process that leads into the perception of covariation between group membership and type of behaviour, we divided the responses on the rating task in two halves. Both sets consist of eight A+, four A−, four B+ and two B− items (thus there is no correlation between group and behaviour). The means for each set are also reported in Table 1. In the first half of the rating task participants seemed to have no clear idea about the groups, which they expressed in a nearly neutral evaluation. Gradually, they indicated differences between the groups in the second half of the rating task with a more positive impression of group A than group B (ic-rate order: $t(34)=4.49$, $p<0.001$, rate-ic order: $t(33)=6.00$, $p<0.001$).

We argued that participants *with prior expectations* can test their expectation against the data in the rating task. The effect of the disconfirming evidence (i.e. positive group B behaviours and negative group A behaviours) seems to be supported by the overall neutral (non-significant) score on the rating task (see Table 1). The violated expectation concerning negative behaviours in group B is clearly expressed in the first half of the statements which is in the opposite direction to the expectation: group A is judged more negatively than group B and this result differed significantly from that of participants without any expectation about the groups: $F(1,148)=12.66$, $p<0.01$. In the second half of the statements group A is favoured, but this effect tends to be weaker for participants who expected negative group B behaviours than for participants without an expectation: $F(1,148)=3.42$, $p<0.07$.

Relation between the rating task and the illusory correlation measures

Berndsen (1994) showed that the responses on the rating task were positively related to the obtained illusory correlation effects. This relationship between the scores on the rating task and the illusory correlation effects suggests that the rating task could serve as another measure of illusory correlation. In the condition where participants have no expectation and started with the rating task, there are significant correlations between the rating index and the scores on the illusory correlation measures. The degree of illusory correlation is positively related to the differential ratings of the groups for this condition: $r=0.51$, $p<0.01$ (evaluation task); $r=0.43$, $p<0.05$ (estimation task); $r=0.67$, $p<0.01$ (assignment task). These findings replicate those of Berndsen (1994). In the other conditions no significant correlations occurred between the rating index and the scores on the illusory correlation measures.

DISCUSSION

The present research focuses on illusory correlation effects as a function of expectations and empirical evidence., Our hypothesis that illusory correlation decreases if the expectation about the negative behaviours in group B is contradicted by empirical data, is supported. When participants have the opportunity to compare their expectation with diagnostic information before they have to judge the degree of covariation, they reported lower illusory correlations than participants who were not first able to test their expectation. Moreover, the relative weight of data-based processes is clearly illustrated in the first half of the rating task where the obviously strong expectation is *disconfirmed.* In other words, covariation judgments are based on an interplay between prior expectation and actual data. This conclusion supports Alloy and Tabachnik's (1984) and McGarty *et al.*'s (1993) viewpoints, rather than that of Nisbett and Ross (1980) and Crocker (1981, 1982) who suggested that covariation judgments are primarily based on expectancies when these are inconsistent with empirical data.

What is more, the findings also support our hypothesis that the rating task will induce illusory correlation effects when there is no prior expectation. Fiedler *et al.* (1984) demonstrated that the expectation-based illusory correlation occurs during the encoding stage of stimulus information. The point to make here is that *data-based illusory correlation* also occurs during the encoding stage. More specifically, participants without expectations reported covariations in the rating task as well as in the illusory correlation measures, but only when the rating task was first. The process leading to the perception of illusory correlation is revealed by inspection of the scores on the rating task. That is, participants did not differentiate between the groups at the beginning of the rating task, but in the second half of the task they seemed to develop the idea, or hypothesis, that the groups differ. The data in the rating task offer the opportunity to *confirm* this hypothesis resulting in stronger illusory correlation effects on the subsequent illusory correlation measures than those of participants who completed the tasks in the reverse order.

Where participants with no prior expectations completed the rating task before the illusory correlation measures there were strong correlations between the illusory correlation measures and the rating task. This replicates the findings of Berndsen (1994). The absence of significant correlations between those tasks in the condition with the reverse order (and without prior expectations) is fully comprehensible, because that task order provided no basis to differentiate between the groups. Furthermore, no significant correlations occurred between the overall rating index and the illusory correlation effects when participants expected negative behaviours in group B. A plausible explanation for this finding is that the rating task provided information about the validity of their expectations and in particular served to violate those expectations, thus reducing the correlation.

A number of different explanations might account for our results. The illusory correlation in the no prior expectation-rating task first condition might be explained by the distinctiveness explanation of Hamilton and Gifford (1976) because the rating task itself provides an opportunity for the participant to perceive a distinctiveness-based illusory correlation. That is, the co-occurrent infrequent information in the rating task is distinctive, implying that it will be better encoded, and the retrieval of this information will be facilitated. This

results in the observed illusory correlation effects. According to Hamilton such an illusory correlation effect is a product of a memory-based process and he and his colleagues have argued recently (e.g. McConnell, Sherman & Hamilton, 1994a) that on-line processing in the illusory correlation paradigm should eliminate this bias. However, the results on our rating task showed that this task in itself can produce illusory correlations and these are not based on memory but rather on on-line processing. In other words, although the distinctiveness explanation is not ruled out here, it seems that illusory correlation is not necessarily a product of memory-based processing. Moreover, distinctiveness does not explain illusory correlation in the prior expectation conditions or why illusory correlation in these conditions is differentially affected by task order.

Another explanation for our results in the conditions with prior expectations, concerns demand characteristics; these might be responsible for the observed covariations in the conditions where a corresponding expectation exists. Indeed there are some similarities in process between what are conventionally termed demand characteristics and the expectations we have created here. However, a critical aspect of our argument is that such expectations are routinely created by the illusory correlation stimulus situation. The difference here is that we have deliberately manipulated them.

Having said that, if the negative behaviours attributed to group B were simply a result of demand characteristics, one would expect no significant differences in illusory correlation effects between the different task orders. However, we observed lower illusory correlations when the rating task preceded the illusory correlation measures than in the reverse order. Moreover, if demand characteristics have caused the judged covariations, the rating task should also reveal the corresponding negative evaluations of group B, which is clearly not the case in the present experiment where the overall evaluation was rather neutral.

The fit principle proposed as an explanation by McGarty *et al.* (1993) offers a suitable framework to explain the obtained effects in the present experiment. The conditions without prior expectations in the present study are *conceptually* similar to the 'no-information conditions' of McGarty *et al.* (1993). More specifically, participants in their first study were simply exposed to the standard illusory correlation measures without receiving the standard stimulus information. Participants displayed no illusory correlation because they had no basis to differentiate between the stimulus groups. This condition is similar to our condition where participants without expectations (and without stimulus information), completed the illusory correlation measures first. These participants displayed zero illusory correlation.

Our condition where participants had no prior expectation and performed the rating task first is conceptually similar to the no-information condition in Study 2 of McGarty *et al.* In this second study, their participants were given a list of positive and negative behaviours without linking these to the stimulus groups. Thus, before starting the tasks the participants could perceive that there were more positive than negative behaviours. Given this and the knowledge (from the standard instructions) that group A is larger than group B, the participants were able to identify, or rather infer, the fit between group size and desirability of behaviours such that the large group (A) is associated with positive (majority) behaviours. This then resulted in an illusory correlation effect that also occurred in our corresponding condition.

Although the rating task and McGarty *et al.*'s presented list of behaviours differ in their methodological instantiation, they have in common that they allow participants to assess a fit between group size and the desirability of behaviours.

Despite this, there seems to be a contradiction involving the origin of illusory correlation. McGarty *et al.* demonstrated that their observed illusory correlation effects could not result from encoding-based mechanisms (as there was nothing to encode), while we have argued that data-based illusory correlation also occurs during the encoding stage. However, this contradiction is more apparent than real, because both our results, and those of Berndsen (1994), show that the meaning of the statements is changed as the participants strive to differentiate meaningfully between the stimulus groups. The clearest evidence of this is that differentiation between the groups increases in the second half of the rating task.

This process whereby participants develop and modify illusory correlations fits in very well with self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987), which views categorization as an adaptive, sense-making process. Moreover, this sense-making process is also clearly illustrated in the conditions where group B is expected to behave negatively. Having prior expectations about negative behaviours in group B is equivalent to already knowing the fit between group membership and behaviour. It seems that these participants test their expectation (or prior fit) against the data in the rating task. The lack of observed fit resulted in relatively low illusory correlations. Thus, the active process of updating information with regard to the existing expectation offers support to self-categorization theory because participants actively made sense of the situation by attuning their judgments to their expectation as well as to the empirical evidence. In this way, their judgements became categorically meaningful. Current research (Berndsen, Spears, Van der Pligt, & McGarty, in preparation) further examines the validity of this sense-making process and the fit principle.

An interesting point involving the expectation of negative behaviour in group B is that it seems that participants disconfirmed their conceptions of the stimulus groups in the rating task. This finding seems to contradict research on the confirmation bias. Early research stated that people were not inclined to test their hypothesis or idea against facts that might disconfirm their hypothesis, but rather prefer facts that offer confirmation (Wason, 1960). However, more recent research dealing with concrete situations rather than abstract task content revealed that people would falsify an hypothesis when the content of that hypothesis focuses on violations or falsifications of the rule (e.g., Cosmides, 1989; Manktelow & Over, 1990). Both requirements (i.e. concrete task content and attention to disconfirmation of that content) seemed to be fulfilled in the present study; the task dealt with unambiguous positive and negative behaviours of two groups and the hypothesis, or expectation, that one of the groups behaves relatively negatively might increase the salience of exceptions to this rule in the rating task. Therefore, testing this expectation against the data in the rating task resulted in a rejection of that expectation, particularly at the start of the task.

The overall rejection of prior expectations in the rating task in our experiment is in contrast to Fiedler *et al.* (1984) who found an illusory correlation effect corresponding with the prior expectation on that task. A plausible explanation for these different results is that expectations about behaviours might be stronger for meaningful groups (or persons, as in Fiedler *et al.*'s study), than for meaningless

groups (as in our study). In other words, the stronger the belief that an expectation is true, the more one might search for confirmation (*cf.* Swann & Giuliano, 1987).

This brings us also to another observation involving the differential learning effects of the rating task. That is, in the absence of prior expectations, participants began to form an impression of the groups in the second half of the rating task such that group A was more favoured than group B. In other words, these participants seem to evaluate the whole data set in the rating task. In contrast, for those participants who expected negative behaviour in group B testing their expectation against the data at the start of the rating task was sufficient to disconfirm this expectation. Thus assessing whether there is a relation between group and behaviour seems to take longer than rejecting an existing expectation involving this relation. However, for these participants, the data in the second half of the rating task may have reconfirmed the original expectation about the negative behaviours in group B. This then resulted in illusory correlation which would be in line with Fiedler *et al.*'s (1984) suggestion that expectation-based illusory correlation is (eventually) effective in the encoding phase. The observed illusory correlation effects are as strong (but not significantly stronger) as those of participants without expectations who completed the tasks in the same order (i.e. rating task followed by illusory correlation measures). A possible explanation for this finding is a contrast effect. After an initial phase of disconfirmation, participants might become aware of the number of cases supporting their original expectations in the second half of the rating task. This could lead to an overcorrection of those expectations resulting in subsequent strong illusory correlation effects.

This process also illustrates that people may not employ prior expectations as a deep-rooted conviction but rather as an hypothesis—the validity of which is tested against empirical data. Of course, there will be limits to such processes of hypothesis testing and evaluation of the evidence. For example, when people strongly believe that their expectation is true, as we argued before, or when they have some interest in maintaining their expectation they may mainly search for confirmation (*cf.* Berndsen, Spears & Van der Pligt, in press). However, such an interest, or involvement, does not play a role in the present study. Therefore, the active hypothesis-testing behaviour indicates that people in the illusory correlation paradigm might search for meaningful intergroup differentiation and this suggests that illusory correlation is not a product of distortion but rather the result of meaning enhancement.

Finally, it is worth commenting on the implications of the innovative methodological features of this experiment. Our results invite the interpretation, as do previous ones (Haslam *et al.*, 1996; McGarty *et al.*, 1993), that standard features of an experimental paradigm (in this case the illusory correlation paradigm) create expectations which although not those originally envisaged by the researchers who devised the paradigm, are nevertheless critical to the social psychological reality of the task (for a discussion of some related points see Bless, Strack & Schwarz, 1993). The results presented here extend this principle in one important respect in that they demonstrate not merely the trite point that the order of presentation of dependent variables affects responses to those variables, but the far more significant observation that for the researcher to ask questions of research participants also involves communicating information to those participants. Under conditions where participants already have, or are in the process of developing, expectancies about the

task and its importance, we should expect (as we have demonstrated here) that responses will be profoundly changed in ways which are crucial to the process under observation.

REFERENCES

Alloy, L. B., & Tabachnik, N. (1984). Assessment of covariation by humans and animals: The joint influence of prior expectations and current situational information. *Psychological Review*, **91**, 112–149.

Berndsen, M. (1994). Meanings inferred by subjects in the illusory correlation paradigm. Unpublished manuscript. Amsterdam: Universiteit van Amsterdam, Vakgroep Sociale Psychologie van Faculteit Psychologie.

Berndsen, M., Spears, R., & Van der Pligt, J. (in press). Illusory correlation and attitude-based vested interest. *European Journal of Social Psychology*, **25**.

Berndsen, M., Spears, R., Van der Pligt, J., & McGarty, C. (manuscript in preparation). Differentiation and illusory correlation.

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.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behaviour*, **6**, 151–155.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, **31**, 187–276.

Crocker, J. (1981). Judgment of covariation by social perceivers. *Psychological Bulletin*, **90**, 272–292.

Crocker, J. (1982). Biased questions in judgment of covariation studies. *Personality and Social Psychology Bulletin*, **8**, 214–220.

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.

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., & Rose, T. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology*, **39**, 832–845.

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.

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

Manktelow, K. I., & Over, D. E. (1990). Deontic thought and the selection task. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds), *Lines of thinking: reflections on the psychology of thought*. Chichester: Wiley.

McArthur, L. Z., & Friedman, S. A. (1980). Illusory correlation in impression formation: Variations in the shared distinctiveness effect as a function of the distinctive person's age, race, and sex. *Journal of Personality and Social Psychology*, **39**, 615–624.

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994a). On-line and memory-based aspects of individual and group target judgments. *Journal of Personality and Social Psychology*, **67**, 173–185.

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994b). 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.

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

Nisbett, R. E., & Ross, L. (1980). *Human inference: strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

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.

Schaller, M., & Maass, A. (1989). Illusory correlation and social categorization: Towards an integration of motivational and cognitive factors in stereotype formation. *Journal of Personality and Social Psychology*, **56**, 709–721.

Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology*, **27**, 107–123.

Snyder, M., & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology*, **14**, 148–162.

Spears, R., Eiser, J. R., & Van der Pligt, J. (1987). Further evidence for expectation-based illusory correlations. *European Journal of Social Psychology*, **17**, 253–258.

Swann, W. B., & Giuliano, T. (1987). Confirmatory search strategies in social interaction: How, when, why and with what consequences? *Journal of Social and Clinical Psychology*, **5**, 511–524.

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.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, **12**, 129–140.