Consciousness and Cognition 63 (2018) 11–28

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Full Length Article

Illusory correlations despite equated category frequencies: A test of the information loss account

Michael Weigl[a,*], Axel Mecklinger[a], Timm Rosburg[a,b]

[a] Department of Psychology, Experimental Neuropsychology Unit, Saarland University, D-66041 Saarbrücken, Germany
[b] Department of Forensic Psychiatry, University Psychiatric Clinics Basel, Wilhelm Klein-Str. 27, CH-4002 Basel, Switzerland

ARTICLE INFO

Keywords:
Illusory correlation
Distinctiveness
Memory
Stereotyping
Availability

ABSTRACT

Illusory correlations (IC) are the perception of covariation, where none exists. For example, people associate majorities with frequent behavior and minorities with infrequent behavior even in the absence of such an association. According to the information loss account, ICs result from greater fading of infrequent group-behavior combinations in memory. We conducted computer simulations based on this account which showed that ICs are expected under standard conditions with skewed category frequencies (i.e. 2:1 ratio for positive and negative descriptions), but not under conditions with equated category frequencies (i.e. 1:1 ratio for positive and negative descriptions). Contrary to these simulations, our behavioral experiments revealed an IC under both conditions, which did not decrease over time. Thus, information loss alone is not sufficient as an explanation for the formation of ICs. These results imply that negative items contribute to ICs not only due to their infrequency, but also due to their emotional salience.

1. Introduction

The ability to extract regularities from a limited number of observations is one of the most fundamental tools an organism needs for adaptive behavior. Humans are even able to implicitly learn complex artificial grammars (e.g. Reber, 1967). In fact, humans have such a strong propensity to detect patterns that they perceive contingencies in environments, even when there are no contingencies – a phenomenon called illusory correlation (Chapman, 1967; Hamilton & Gifford, 1976).

More technically, an illusory correlation (IC) is a subjectively perceived correlation between two events, which differs systematically from the actual covariation between those events (e.g. Chapman, 1967; Fiedler, 2000). The two events might actually not correlate at all or correlate in the opposite direction as reported. ICs have been investigated in basic research (e.g. Chapman, 1967; Tversky & Kahneman, 1973) as well as in applied research, like psychodiagnostics (Chapman & Chapman, 1967), clinical psychology (Alloy & Abramson, 1979), or organizational psychology (Feldman, Camburn, & Gatti, 1986). Very fruitful investigations on the IC were conducted in stereotyping research (e.g. see Hamilton, 1981 or Stroessner & Plaks, 2001 for reviews).

There are (at least) two different types of ICs discussed in the stereotyping literature – each associated with a specific pattern of results (Hamilton, 1981; Tversky & Kahneman, 1973; see Fiedler, 2000, for a more fine-grained classification): the expectancy-based and the distinctiveness-based illusory correlations. In expectancy-based ICs, participants already have an expectation about the relationship of two variables, based on their experiences and personal beliefs. When study participants have to judge the covariation between well-known groups (e.g. accountants and salesmen) and certain traits (e.g. timid and talkative) in a new set of stimuli, their

---

[*] Corresponding author at: Department of Psychology, Experimental Neuropsychology Unit, Saarland University, Campus, Building A2.4, D-66123 Saarbrücken, Germany.

E-mail address: m.weigl@mx.uni-saarland.de (M. Weigl).

https://doi.org/10.1016/j.concog.2018.06.002
Received 26 September 2017; Received in revised form 28 April 2018; Accepted 3 June 2018

judgment on the new set is usually consistent with their pre-experimental expectations (Hamilton & Rose, 1980).

In contrast, participants in experiments on the distinctiveness-based IC have to infer a correlation about material for which they do not possess preexisting expectations about the relationship. In the framework of the current study, we primarily refer to the distinctiveness-based IC.

In the seminal study of Hamilton and Gifford (1976), participants read short descriptions about members of two fictional groups – group A and group B, with group A having twice as many members as group B. For both groups, two-thirds of the description referred to desirable behavior and one-third to undesirable behavior. In other words, group membership and behavior were uncorrelated. Nevertheless, the participants showed a tendency to associate the majority with the frequent, desirable behavior and the minority with the infrequent, undesirable behavior. This pattern was observed consistently across a range of dependent measures (evaluative trait ratings, frequency judgments, and cued recall performance; Hamilton & Gifford, 1976; see Mullen & Johnson, 1990 for a review). The popularity of the concept of ICs stems from the fact that it offers a cognitive explanation for the formation of stereotypes. Moreover, the experimental set-up resembles the situation we encounter in our modern societies: minorities are by definition smaller than majorities and most people behave in a norm-consistent, desirable way (e.g. Alves, Koch, & Unkelbach, 2017a; Fiske, 1980; Kanouse, 1984).

Skewed frequency distributions are assumed to be essential for ICs to arise (e.g. Fiedler, 1991, 1996; Hamilton & Gifford, 1976). In cases in which undesirable behavior is more frequent than desirable behavior, the IC is reversed, i.e. the majority is evaluated less favorable than the minority (Hamilton & Gifford, 1976, Exp. 2; Mullen & Johnson, 1990). However, there is a still ongoing debate about the mechanisms by which skewed frequency distributions influence our judgment and a variety of models have been put forward to explain illusionary correlation. As a consequence, ICs have been investigated from various theoretical perspectives (e.g. accentuation approach: McGarty, Haslam, Turner, & Oakes, 1993; availability account: Rothbart (1981); memory trace model: Smith, 1991; pseudocontingencies: Fiedler, Freytag, & Meiser, 2009; recurrent connectionist model: Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003; Rescorla-Wagner model: Murphy, Schmeer, Vallée-Tourangeau, Mondragón, & Hilton, 2011; attention theory: Sherman et al., 2009). Two prominent accounts are the shared distinctiveness account (SDA) and the information loss account (ILA). Our study was designed to test predictions made on the basis of these two accounts. For the sake of clarity, we will describe only the SDA and ILA at this point; details of the other accounts can be found in the General Discussion.

The SDA states that infrequent combinations are more distinctive and, therefore, better encoded than more common ones (Chapman, 1967; Hamilton & Gifford, 1976; Tversky & Kahneman, 1973). Infrequent combinations are therefore more easily available in memory than others. As individuals estimate the frequency of the combinations on the basis of their availability, infrequent combinations are overestimated (Tversky & Kahneman, 1973).

The evaluation of the empirical evidence for the SDA is unfortunately hampered by the fact that researchers conceptualized distinctiveness quite differently. For example, Hamilton and Gifford (1976) defined distinctiveness as infrequency, whereas Feldman et al. (1986) also considered negativity as distinctive.

Schmidt (2012) identified four different types of distinctiveness: (1) primary distinctiveness (distinctiveness in the immediate context, usually due to infrequency), (2) secondary distinctiveness (distinctiveness over life-time, usually due to bizarreness), (3) emotional significance (emotional engaging stimuli) and (4) high priority stimuli (relevant, but non-arousing stimuli). For the purpose of the current study, we define a stimulus as being distinctive, if it fulfills one of these four criteria.

Evidence for better memory for shared distinctive items stems from studies using free recall (Hamilton, Dugan, & Trolier, 1985) and one-shot ICs (Risen, Gilovich, & Dunning, 2007). Further support can be found in the memory literature: Distinctive items are in general better remembered than non-distinctive items (e.g. Alves et al., 2015; Fabiani & Donchin, 1995; Hunt, 1995, 2009; von Restorff, 1933; see also Schmidt, 1991, 2012, for an integrative account). Furthermore, memory is even better for items that are distinctive on several stimulus dimensions (e.g. Hunt & Mitchell, 1982; Kuhbandner & Pekrun, 2013).

The information loss account (ILA) offers an alternative explanation for ICs without assuming any differential processing of information (Fiedler, 1991, 1996, 2000; Smith, 1991). Since perception and memory are far from perfect, noise can distort parts of information during encoding, storage, and retrieval. For example, due to such noise, we might misremember a rude person from the majority as a member of the minority. If participants are asked to make a judgment about their attitude towards the groups, unbiased aggregation of these distorted data alone is sufficient to lead to the erroneous conclusion that a correlation between groups and behavior is present. In the typical IC experiments the distribution of the valence of the behavior is skewed, i.e. positive stimuli are objectively more frequent than negative stimuli (or vice versa; Fig. 1). For the majority, the preponderance of positive (or negative) behaviors becomes evident to the participant during encoding, because they can aggregate over a large number of instances. Therefore, the subjective frequency estimates for the majority are not so much affected by noise and should roughly correspond to the actual frequencies. The minority, however, is more strongly affected by this noise, because single outliers have more influence on the estimates of smaller samples and the estimates regress to the mean (see 33% and 67% condition in Fig. 1). Thus, the main appeal of the ILA is its parsimony. ICs can be explained without assuming biased processing (e.g. different processing of distinctive and non-distinctive information).

Evidence for the ILA stems from computer simulations that reproduce the IC effect without the assumption of biased processing (Fiedler, 1996, 2000; Smith, 1991). But there is also experimental evidence that the overestimation of frequencies increases when categories are split into sub-categories (Fiedler, 1991; Fiedler & Armbruster, 1994) or that ICs can be observed even in the absence of distinctive or infrequent information (Fiedler, 1991; Shavitt, Sanbonmatsu, Smittipatana, & Posavac, 1999; Van Rooy, Vanhoomissen, & Van Overwalle, 2013).

The SDA and the ILA are not mutually exclusive. However, it is necessary to test boundary conditions to judge the relative merit of both accounts: One highly important boundary condition arises naturally from a closer look at the regression to the mean argument of

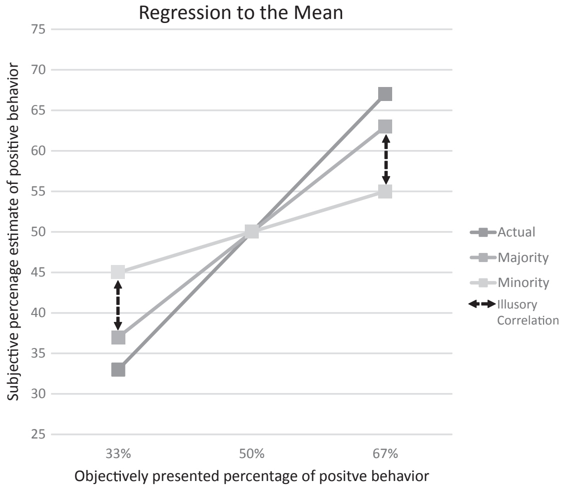

Fig. 1. Regression to the mean as a potential cause for illusory correlation. In the 67% condition, positive behavior is presented twice as often as negative behavior. In the 33% condition, in contrast, negative behavior is presented twice as often as positive behavior. In both cases, the subjective estimates of the majority (dark grey line) are closer to the actually presented frequency (black line) than the estimates of the minority (light grey line). The illusory correlation manifests itself as the difference between the subjective estimate of positive behavior for the majority and the minority (dotted double arrow). Please note that no illusory correlation would be expected if positive and negative behavior were equally frequent (50% condition). Adapted from Fiedler and Walther (2004, Figure 3.2).

the ILA. An illusory correlation would be expected only if the frequency distribution for behavior is skewed (33% and 67% condition in Fig. 1), because only these circumstances allow differences in regression between the majority and the minority. Thus, no IC would be expected when positive and negative behaviors are equally distributed (50% condition in Fig. 1).

A reassessment of the classical IC paradigm reveals that several kinds of distinctiveness are involved: Minority members are distinctive in the context of presentation due to infrequency (primary distinctiveness). Negative behavior is not only distinctive due to rarity in the stimulus set (primary distinctiveness), but also per se due to its rarity in real life (secondary distinctiveness) and unpleasantness (emotional significance) as documented by the literature on the asymmetry between positive and negative information (Fiske, 1980; Kanouse, 1984). Furthermore, positive information is more similar than negative information which additionally renders negative information distinctive (Alves, Koch, & Unkelbach, 2016, 2017b; Koch, Alves, Krüger, & Unkelbach, 2016). Thus, equating the category frequencies for negative and positive stimuli allows dissociating primary distinctiveness (resulting from infrequency) from other forms of distinctiveness.

In such a setting with equated frequencies, the SDA would still predict the presence of an IC, whereas the ILA would predict its absence, because regression to the mean would equally affect positive and negative items, as illustrated (Fig. 1).

The first and most important research question of the present study was, thus, whether an IC could be observed in a condition with equated frequencies (with positive and negative items being equally frequent). For this purpose, findings in the equated frequency condition were compared to a condition with skewed frequencies (i.e. the standard paradigm with positive items twice as frequent as negative items). To the best of our knowledge, no previous study has directly addressed such a comparison.

A second research question concerned the relationship between memory and IC. Both, the SDA and the ILA, assume a causal contribution of memory in the formation of ICs, even though they assign different roles to memory. Based on the SDA, memory should be better for the minority than for the majority and best for negative items of the minority. Moreover, the better memory for such distinct group-behavior combinations should be predictive for the extent of ICs. In contrast, the ILA would predict similar memory performance for all group-behavior combinations, because the random noise is supposed to affect all of them in the same manner.

The third research question of the present study was whether ICs would remain stable over time – a question still understudied. Feldman et al. (1986) investigated ICs using delays of 12 or 24 h. Unfortunately, the results of this study are inconclusive, because their paradigm failed to induce a reliable IC in the first place. Distinctiveness effects on episodic memory are relatively stable over time (e.g. Hunt, 2009). Thus, on the basis of the SDA, we would expect the IC to be unaffected by a time delay. However, it also seems plausible to assume that the distortion of the encoded material due to information loss increases over time. Memory steeply declines over a short range of time after initial encoding (Ebbinghaus, 1885/1966; see Rubin & Wenzel, 1996 for an extensive review). Thus, introducing a short delay after initial encoding should already lead to some information loss and regression to the mean. The extent of IC as a function of noise presumably follows an inverted u-shaped pattern. As long as estimates for both, the majority and the minority, can still regress, increases in noise will increase the extent of IC, because more information will be lost for the minority than the majority. If information about the minority is almost completely lost and only the estimates for the majority can show further regression, then increases in noise will reduce the difference between both groups and the extend of IC (see also Fiedler, Russer, & Gramm, 1993). Thus, while the SDA predicts that a time delay does not affect ICs, the ILA predicts that a short time delay after establishing an IC might lead to an increase of ICs.

As a first step, we tested our theoretical assumptions on the impact of equated category frequencies and temporal delay in a

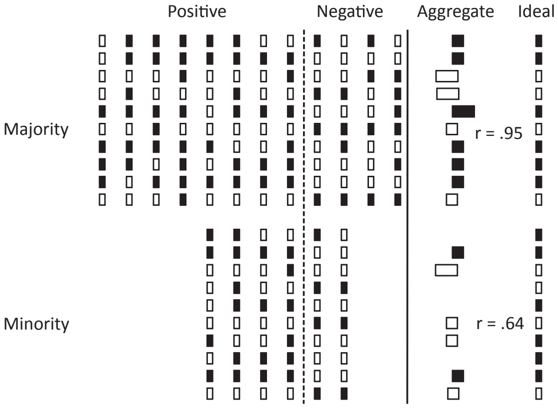

Fig. 2. Illustration of the aggregation process in BIAS. The majority is represented by 12 stimuli – 8 positive and 4 negative. The minority is represented by 6 stimuli – 4 positive and 2 negative. During the aggregation process, each row is summed up. Black rectangles are counted as +1 and white rectangles as -1. The aggregate vector is then correlated with the ideal type of positive behavior. The aggregate of the majority correlates more closely with the ideal type than the aggregate of the minority. Adapted from Fiedler (1996, Fig. 2).

computer simulation based on the Brunswikian Induction Algorithm for Social cognition (BIAS; Fiedler, 1996, 2000; Fiedler & Walther, 2004). BIAS implements the assumptions of the ILA that the mere unbiased processing of information in a noisy environment is sufficient for the formation of ICs. Subsequently, we tested the predictions derived from the simulation in two behavioral experiments – one with the standard skewed distribution (Experiment 1) and the second with an equal distribution (Experiment 2).

2. Simulation study

2.1. Method

BIAS applies the principles of information loss and aggregation on stimulus matrices to simulate cognitive biases. Computationally, the BIAS model assumes that information is represented in a stimulus matrix in which the columns represent individual stimuli or events and the rows represent cues (e.g. features of a stimulus), i.e. a stimulus is not a scalar value, but a vector (Fig. 2; Fiedler, 1996, 2000). Each stimulus vector is derived from the distal entity. The distal entity is a vector that defines the true values, i.e. how the values of the stimulus would look like in an ideal, noise-free environment. Due to noise (e.g. misperceptions, forgetting, natural variations) the stimulus vector will deviate from the distal entity, i.e. some elements of the stimulus vector randomly differ from the distal entity. For example, red and white roses are both instances of the distal entity roses and, therefore, resemble each other even though genetic and ontogenetic variations alter their appearance to some degree. The error variance due to noise is removed by applying an aggregation rule (e.g. averaging or summation) on each row. If there is a prevailing tendency in the data, it will become apparent after aggregation. If, for example, a person is asked about his/her attitude towards a product, this person will aggregate over arguments for and against the product. The attitude would be positive, if there are more pros than cons (Fiedler, 1996).

In our simulation, a stimulus was represented by a vector with ten elements (Fig. 2; Fiedler, 1996). Each element represented a feature or cue. The presence of a feature was coded as 1, and the absence of a feature was coded as -1. The ten elements represented the valence of the trait. Positive and negative traits were coded as the exact opposite of each other. In order to explore the impact of the noise level on the extent of IC, either one, two, three, four, or five elements (corresponding to total information loss) were randomly chosen and reversed.

For the simulation of the skewed frequency condition, the simulation included 16 positive and 8 negative behaviors for the majority, as well as 8 positive and 4 negative behaviors for the minority. For the simulation of the equal frequency condition, the simulation included 12 positive and 12 negative behaviors for the majority, as well as 6 positive and 6 negative behaviors for the minority. In order to assess the prevailing tendency for each run, the stimuli were summed up row-wise for each group separately and the resulting aggregated vector was correlated with the predefined ideal vector for positive stimuli (Fig. 2). A high correlation means that the aggregate corresponds closely to the ideal vector. The average correlations scores of the two groups across 1000 simulations were compared by paired t-tests. In these simulations, an illusory correlation was judged to be present, if the aggregate of the majority correlated significantly higher with the ideal vector of positive stimuli than the aggregate of the minority. All simulations were run in R 3.3.1 (R Core Team, 2016) using custom-written R code.

2.2. Results

The results of the computer simulation are depicted in Fig. 3. As hypothesized, there were ICs in the skewed frequency condition for the noise levels one to four, i.e. the correlation between the ideal vector for positive behavior and the aggregate vector was higher for the majority than for the minority (all t-values > 21.03, $p < .001$). Thus, the overall positivity became more apparent for the

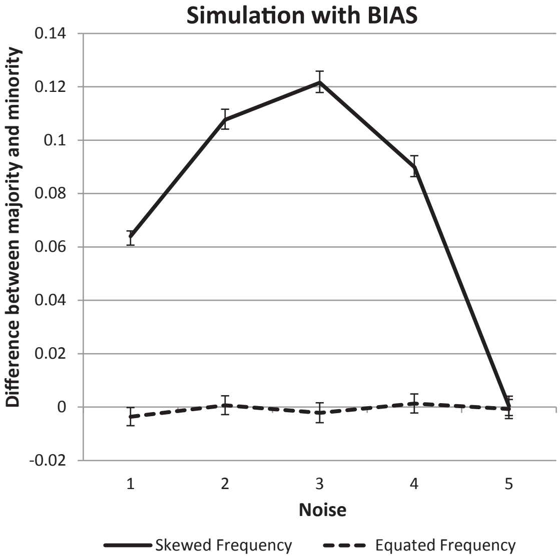

Fig. 3. Results of the computer simulation with BIAS. The illusory correlation is depicted as difference between the majority's correlation with the ideal vector and the minority's correlation with the ideal vector for the skewed frequency condition (solid line) and the equated frequency condition (dashed line). In the skewed frequency condition an illusory correlation was present at all noise levels but the 5th. In the equated frequency condition, the illusory correlation was absent at all noise levels.

majority than the minority at these noise levels. As expected, no IC was observed at the highest noise level when all five elements were changed ($t(999) = 0.10, p = .917$), because all information was erased by noise. Also as predicted, the extent of ICs initially rose with increasing noise levels and then fell for more extreme levels of noise: The ICs of all five noise levels significantly differed from each other (all p-values $< .008$). In contrast, for the equated frequency condition, there were no significant ICs at any noise level ($|t(999)| < 1.06, p > .289$) and no differences between the five noise levels either (all p-values $> .330$).

To sum up, the simulations showed that ICs were only present in the skewed frequency condition and the extent of ICs initially increased with increasing noise and then declined with extensive noise levels. These results have two implications for our experiments: (1) If nothing but information loss exerts an effect on judgment, an IC should be observed in Experiment 1 (skewed frequency condition), but not in Experiment 2 (equated frequency condition). (2) If information loss is moderate, a short delay should increase the extent of IC. Thus, if the assumptions of the ILA hold, the pattern of the behavioral data should resemble the pattern found in the simulations.

3. Experiment 1: Skewed frequency condition

3.1. Method

3.1.1. Participants

Thirty-four students (26 female) of the Saarland University participated in Experiment 1. Participants received course credit or comparable compensations for their participation. Data of six participants were not included in the analysis. One participant did not use the correct response keys. The five other excluded participants were either not German native speakers ($n = 2$), who were granted participation in the study in order to obtain course credits, or were participants who inferred the hypotheses of the study as reported in the post-experimental questionnaire ($n = 3$). The latter three participants reported quite specific hypotheses about the purpose of the experiment (e.g. that the experiment was about distinctive features of minorities or about the development of stereotypes about minorities, or even explicitly mentioned illusory correlations) and were excluded on the grounds of previous findings: The awareness of influence can hamper the investigation of judgmental biases (see Bless, Fiedler, & Strack, 2004, pp. 122–124, for a discussion). Specifically to the illusory correlation, knowledge about the task has been shown to reduce the extent of illusory correlation (e.g. Chapman, 1967; Lilli & Rehm, 1983, 1984). Thus, from our point of view, the inclusion of these participants would diminish the validity of the data. The exclusion of the three participants did, however, not affect the major findings. The final sample comprised 28 participants (23 female, median age 23 years, range 18–31 years).

3.1.2. Materials

For Experiment 1, 24 positive and 12 negative adjectives were drawn from a pool of 48 trait adjectives and matched for arousal, imageability, and word length (see Supplementary Materials for details). Descriptions of traits were chosen instead of descriptions of behavior, because by this approach sentence lengths and, thus, encoding times could be equated. Thirty-six German male first names that were popular in the time period from 1986 to 1993 were selected from a website (http://www.beliebte-vornamen.de/). We

Table 1
Distribution of positive and negative traits across groups for Experiment 1 and 2.

	Experiment 1		Experiment 2	
	Positive traits	Negative traits	Positive traits	Negative traits
Majority	16	8	12	12
Minority	8	4	6	6

selected this time period to ensure that the participants were familiar with the first names. Consistent with other studies on the IC (e.g. Hamilton & Gifford, 1976; Stroessner, Hamilton, & Mackie, 1992), the names were restricted to male names in order to control for possible effects of the stimulus persons' gender.

For each participant, a list of 36 person descriptions was created. The 36 first names and trait adjectives were randomly combined to a description and assigned either to the majority, group A, or to the minority, group B. As shown in Table 1, sixteen descriptions with positive traits and eight descriptions with negative traits were assigned to group A, eight descriptions with positive traits and four descriptions with negative traits were assigned to group B.

3.1.3. Procedure

The experiment was programmed and run using E-Prime 2.0. Participants sat in front of a 17 in. monitor and were individually tested. All displays were centered and had white background. Words and sentences were presented in black 18pt Courier New font. The instruction followed those of Hamilton and Gifford (1976) and Pryor (1986). The experiment was described as being concerned with how people perceive and retain information about others. Participants were told that they would read descriptions of students made by persons close to them and that each person belonged to one of two groups, which actually existed and were arbitrarily named group A and group B for the purpose of the experiment (see Table 2 for an overview over the experiment).

During encoding participants saw 36 descriptions. They were instructed to read the descriptions and memorize all the information for a subsequent memory test. Each trial started with a fixation cross presented for 500 ms. Then a blank appeared for 200 ms. Next, the descriptions appeared on the screen for 4000 ms. Each description contained a male first name, the group membership, and a positive or negative trait (e.g. "Oliver from group A is nice." or "Andreas from group B is stubborn."). The descriptions were presented in a random order. At the end of the trial another blank was presented for 200 ms.

After the encoding task participants had to count backward from 100 in steps of three for one minute and to enter the number they reached. Then they started the group assignment task. This task was intended to assess the bias against the minority and probe the source memory for group membership. Participants were told to assign the descriptions to one of the groups as fast and as accurately as possible. Each trial started with a blank presented for 200 ms. Next, the sentences from the encoding task were presented again in random order without group information (e.g. "Oliver is nice."). Participants could respond for 2500 ms by pressing the F or the J key on the keyboard. Response keys to group assignments were counterbalanced across participants. Only when participants did not respond within the given time window, a feedback screen appeared for 500 ms, reminding the participant to respond faster on the

Table 2
Overview over the procedure in Experiment 1 and 2. The tasks were presented in the order in which they are listed (except the frequency estimation and the evaluative trait rating for which the order was counterbalanced across subjects). The list below each task summarizes the measures we derived from this task.

Encoding Task
Filler Counting Task (1 min.)
Immediate Testing ● Group Assignment (assessing bias against the minority, illusory correlation, and source memory performance) ● Frequency Estimation (assessing bias against the minority and illusory correlation) ● Evaluative Trait Rating (as in Frequency Estimation)
Filler Task (40 min.)
Delayed Testing ● Group Assignment (assessing bias against the minority, illusory correlation, and source memory performance) ● Frequency Estimation (assessing bias against the minority and illusory correlation) ● Evaluative Trait Rating (as in Frequency Estimation)
Sentence Valence Rating (calculating the encoded correlation)
Control Questionnaire

next trials. Each trial ended with another blank shown for 200 ms.

Next, participants filled out computerized versions of the frequency estimation task, in which participants estimated the percentage of negative traits for each group separately, and the evaluative trait rating, in which participants rated each group separately on ten traits (*helpful*, *tolerant*, *sociable*, *affectionate*, *honest*, *ingenious*, *friendly*, *unreliable*, *industrious*, and *irresponsible* taken from Fiedler et al., 1993) using a 10-point rating scale. The frequency estimation task and the evaluative trait rating as well as the group labels were counterbalanced across participants.

For about 40 min participants conducted an unrelated filler task. After this delay participants performed the group assignment task, the frequency estimation task, and the evaluative trait rating for a second time in the same sequence as in the first part of the experiment. After these tasks, the sentences of the encoding task were presented again. Participants had to rate the valence of each sentence by using the keys 1, 2, and 3, i.e. whether the sentence was negative, neutral, or positive. This sentence valence rating served as a manipulation check to ascertain that the sentences were accurately encoded according to our intended design. When participants pressed one of the keys, a blank was presented for 400 ms followed by the next sentence.

At the end of the experiment the participants were given a questionnaire which asked for participants' sensitivity for the hypotheses and related control questions.

3.1.4. Data analysis

At both, immediate and delayed testing, group assignments, frequency estimation, and evaluative trait rating were used to assess the amount of bias against the minority. In order to control for Type I errors while at the same time preserving statistical power, these measures were subjected to a multivariate analysis of variance (MANOVA) with the factors Group and Time point. We used the difference between the relative frequency of positive and negative descriptions assigned to each group in the group assignment task, the estimated relative frequency of negative behavior for each group, and the mean evaluative trait rating scores for each group as dependent measures. Significant main effects or interactions were followed up with univariate analyses of variance (ANOVA).

Source memory performance was used to test for better memory of shared-distinctive items. Memory was assessed by calculating the unbiased hit rates (Wagner, 1993) for the valence condition (positive vs. negative) and the two time points (immediate vs. delayed) separately. Unbiased hit rates are the conditional probability that the stimulus is correctly classified given the stimulus is shown (e.g. correctly assigning “A” to a group A item) multiplied with the conditional probability that the correct response category is chosen given this response category (e.g. correctly responding “A” when responding “A”).

The unbiased hit rates were transformed with an arcsine transformation (Wagner, 1993) to ensure normal distribution and entered in a repeated measure ANOVA with Group (majority vs. minority), Valence (positive vs. negative), and Time point (immediate vs. delayed) as independent variables.

In order to provide a comprehensive picture of the memory performance, we also analyzed the memory data using Pr, a measure for discrimination, and Br, a measure for response bias (Snodgrass & Corwin, 1988). Pr and Br were calculated for negative and positive valence and for the two time points (immediate and delayed) separately and entered in a repeated measure ANOVA with Valence (positive vs. negative) and Time point (immediate vs. delayed) as independent variables.

In order to quantitatively assess the extent of ICs, correlation coefficients were calculated individually for each participant, dependent variable, and time point separately: Phi coefficients for group and valence were calculated from the group assignment and the frequency estimation data. Point-biserial correlations for group and evaluation were calculated from the evaluative trait rating data from the learning phase and the sentence rating data at the end of the experiment. The latter correlation coefficient represents the encoded correlation. With it, we can not only ensure that participants accurately encoded the sentences according to our experimental design (i.e. that participants interpreted a negative description as negative). We can also preclude that the IC was already formed during encoding (i.e. that erroneous encoding acts as a source of the noise according to the ILA) which might hamper the interpretation of the retrieval data. Please note that the phi coefficient and the point-biserial correlation are both simplified versions of the Pearson correlation coefficient and are therefore comparable with each other. Positive correlations indicate that the majority is associated more with positive descriptions and the minority more with negative descriptions. Fisher z-transformation was applied to all correlation coefficients for statistical tests. All correlation coefficients were subjected to a MANOVA to test for difference from a null vector. If a significant effect was observed, follow-up one-tailed t-tests were conducted.

All statistical analyses were conducted using IBM SPSS 24 (IBM Corp., Armonk, NY, USA). Effect sizes for t-tests were calculated with G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). The alpha criterion was set to $p = .050$ for all analyses. The Benjamini-Hochberg procedure for multiple comparisons was used to adjust the p-values in follow-up tests.

3.2. Results

3.2.1. Group assignments, frequency estimation, and evaluative trait rating

The MANOVA revealed a main effect for Group (Wilk's $\Lambda = 0.64$, $F(3, 25) = 4.64$, $p = .010$). The univariate follow-up analyses indicated that participants assigned more negative descriptions to the minority than to the majority ($F(1, 27) = 14.93$, $p = .003$, $\eta_p^2 = 0.36$; see Fig. 4 Top), estimated the frequency of negative traits to be higher in the minority than in the majority ($F(1, 27) = 4.59$, $p = .041$, $\eta_p^2 = 0.15$; see Fig. 4 Middle), and evaluated the minority less favorable than the majority ($F(1, 27) = 10.30$, $p = .005$, $\eta_p^2 = 0.28$; see Fig. 4 Bottom), even though the ratios of positive and negative descriptions at encoding were equal in both groups. Thus, an IC was present in all three measures. The main effect for Time point was not significant (Wilk's $\Lambda = 0.74$, $F(3, 25) = 2.94$, $p = .052$). Critically, the interaction between Group and Time point was not significant (Wilk's $\Lambda = 0.98$, $F(3, 25) = 0.15$, $p = .930$), indicating that the IC was stable over time.

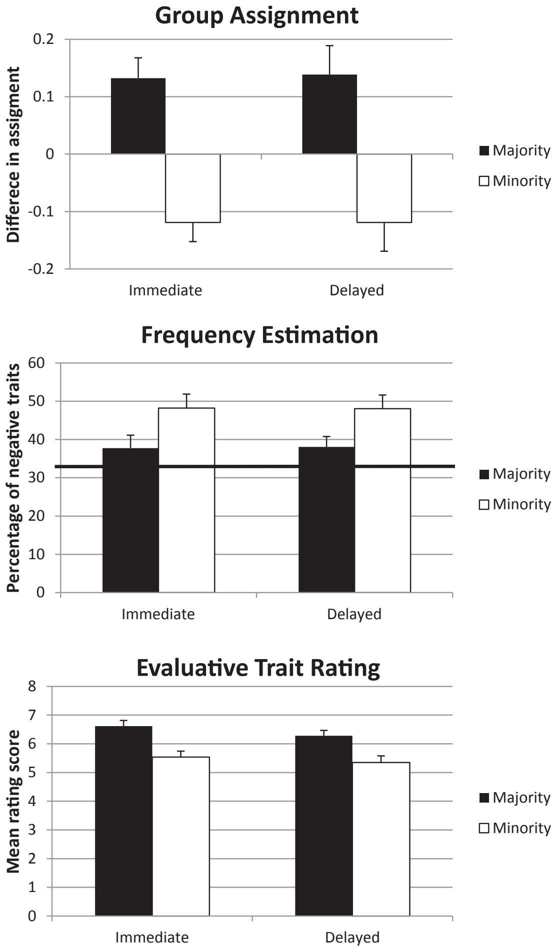

Fig. 4. Overview over the results of Experiment 1. Top: Difference between positive and negative traits assigned to each group. Middle: Estimated ratio of negative trait in percent. The bold line represents the true ratio. Bottom: Mean evaluative trait ratings. Error bars represent between-subject standard errors.

3.2.2. Memory

The results for the unbiased hit rates can be seen in Fig. 5. There was a significant main effect for Group ($F(1, 27) = 147.45$, $p < .001$, $\eta_p^2 = 0.85$). Source memory was better for descriptions of the majority than for descriptions of the minority. This main effect was modulated by Valence ($F(1, 27) = 7.25, p = .012, \eta_p^2 = 0.22$). No other effects were observed (all $Fs < 1.84, p > .187$). The Group × Valence interaction was followed-up by t-tests for descriptions of each group: positive descriptions of the majority were better recalled than negative descriptions of the majority ($t(27) = 3.28$, $p = .006$, two-tailed, Cohen's $d = 0.62$). Contrary to the predictions of the SDA, no difference was observed between the recall of positive and negative descriptions of the minority ($t(27) = -0.67$, $p = .506$, two-tailed, Cohen's $d = -0.13$).

The results for Pr and Br can be seen in Fig. 6. The analysis of item discrimination Pr did not reveal any significant effects (all $Fs < 1$, $p > .332$). In contrast, the analysis of the response bias Br revealed a significant effect for Valence ($F(1, 27) = 8.57$, $p = .007$, $\eta_p^2 = 0.24$) indicating that participants were more likely to assign positive descriptions to the majority than negative descriptions. No other effects were significant (all $Fs < 1, p > .579$).

3.2.3. Measures of illusory correlations

The overall MANOVA was significant (Wilk's $\Lambda = 0.49$, $F(7, 21) = 3.09$, $p = .021$). As can be seen in Table 3, an IC was observed for all correlation coefficients except the coefficient for the frequency estimation task at immediate testing ($p = .053$) and the encoded correlation ($p = .215$). The latter finding indicates that the IC was formed after encoding. Furthermore, the encoded correlation did neither correlate with any single IC coefficient (all p-values $> .282$) nor with the average of all single IC coefficients ($r = -0.07$, $p = .708$).

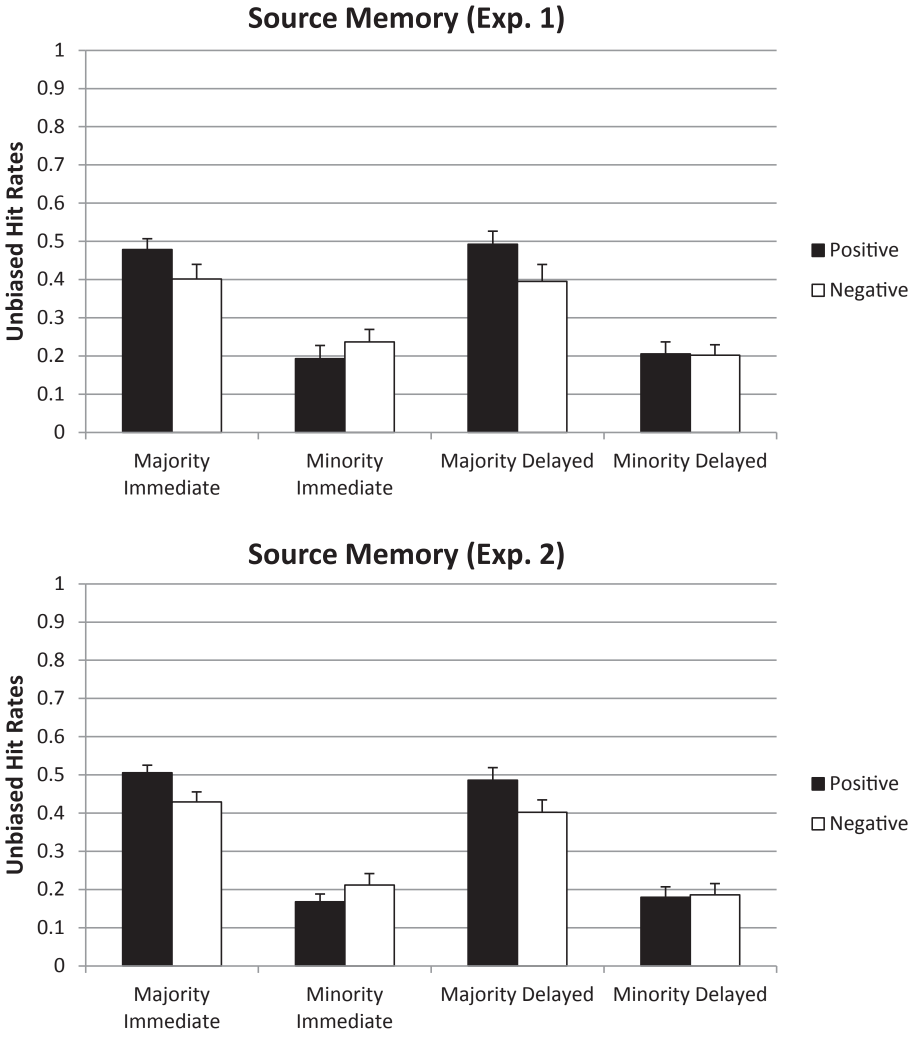

Fig. 5. Unbiased hit rates for source memory in Experiment 1 (Top) and Experiment 2 (Bottom). Error bars represent between-subject standard errors.

3.3. Discussion

The aim of Experiment 1 was to reproduce the effect of an illusory correlation for the standard skewed frequency distributions (Hamilton and Gifford, 1976) and to validate our experimental design and materials. Our experiment indeed successfully evoked ICs, which were observed consistently across all task for assessing ICs (group assignment, frequency estimation, and evaluative trait rating). By obtaining these assessments in Experiment 1, we established a baseline for Experiment 2, in which we sought to test our prediction that, contrary to the predictions of the ILA, equated frequencies for positive and negative traits still result in an IC.

Other noteworthy findings of Experiment 1 were that ICs did not change over time and were not attributable to erroneous encoding. Moreover, the analysis of source memory accuracy revealed that descriptions of the majority were better remembered than descriptions of the minority, and positive descriptions of the majority were better remembered than its negative descriptions. Contrary to the SDA, negative descriptions of group B members were not better remembered than positive descriptions. In addition to this, response bias, but not item discrimination differed between positive and negative descriptions. We review these findings in detail in the general discussion.

4. Experiment 2: Equated frequency condition

4.1. Method

4.1.1. Participants

41 students (35 female) of the Saarland University participated in Experiment 2. Participants received partial course credit or

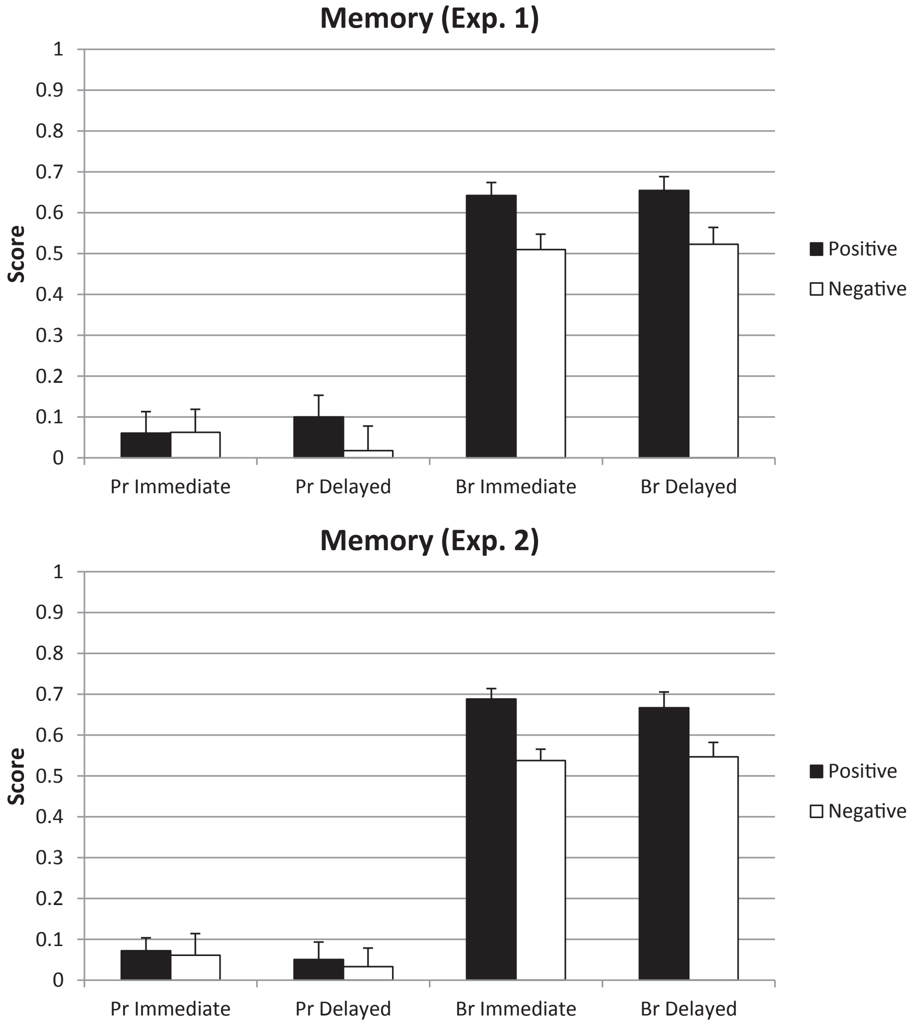

Fig. 6. Pr and Br scores for memory in Experiment 1 (Top) and Experiment 2 (Bottom). Error bars represent between-subject standard errors.

Table 3
Size of the illusory correlation in Experiment 1 and 2 across measures and time points. Please note that the presented measures are untransformed correlation coefficients. Statistical tests, however, were conducted with the Fisher z-transformed coefficients. The Benjamini-Hochberg procedure was used to adjust the *p*-values.

		Group assignment		Frequency estimation		Evaluative trait rating		Encoded correlation	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exp. 1	Immediate	0.12**	0.19	0.11†	0.32	0.26**	0.41	–	–
	Delayed	0.13*	0.27	0.10*	0.27	0.23**	0.35	0.02	0.11
Exp. 2	Immediate	0.13**	0.17	0.10**	0.19	0.25**	0.39	–	–
	Delayed	0.10*	0.30	0.08*	0.22	0.22*	0.43	−0.01	0.07

Significantly different from zero:
† $p < .10$ one-tailed.
* $p < .05$ one-tailed.
** $p < .01$ one-tailed.

comparable compensations for their participation. Those who did not finish the experiment properly ($n = 2$ did not use the correct response keys, $n = 1$ could not finish task due to technical error), who reported severe difficulties to focus on the task ($n = 3$), or who reported quite specific hypotheses about the purpose of the study as reported in the post-experimental questionnaire ($n = 5$) were excluded from further analysis. The exclusion of the participants who inferred the hypotheses of the study rendered some effects

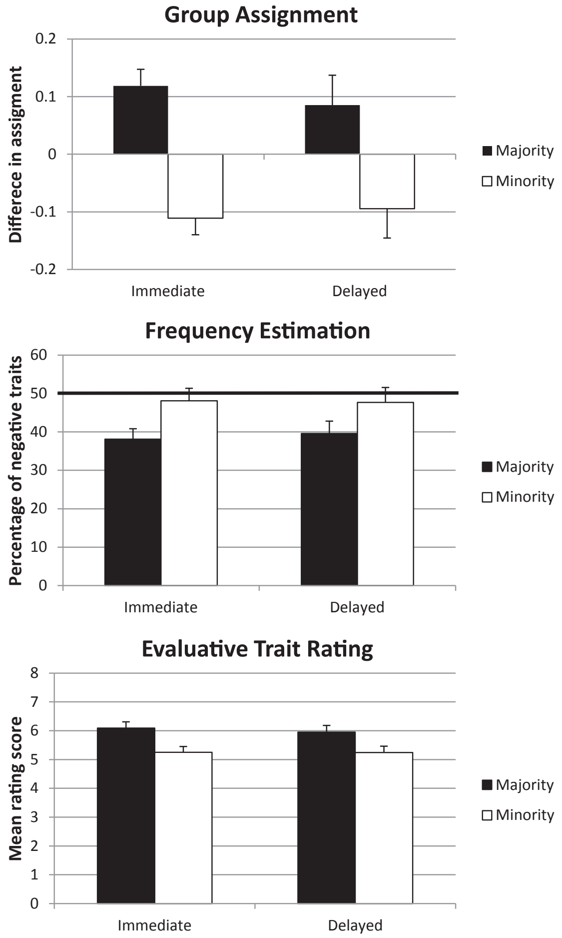

Fig. 7. Overview over the results of Experiment 2. Top: Difference between positive and negative descriptions assigned to each group. Please note that the difference score are smaller than in Experiment 1 due to the equated frequencies of positive and negative descriptions. Middle: Estimated ratio of negative traits in percent. The bold line represents the true ratio. Bottom: Mean evaluative trait ratings. Error bars represent between-subject standard errors.

marginally significant or non-significant. However, similar as for Experiment 1, we opted to exclude the latter participants, because otherwise the validity of the conclusions drawn from the data could be considered as compromised (see Section 3.1.1 for a discussion of the relevant research). The final sample comprised 30 participants (26 female, median age 20 years, range 18–30 years).

4.1.2. *Materials*

For Experiment 2, 18 positive and 18 negative traits were drawn from the pool of 48 adjectives and again matched for arousal, imageability, and word length (see Supplement S1). For each participant, the 36 first names and trait adjectives were randomly combined to a description and assigned either to group A (majority) or to group B (minority). Twelve positive and twelve negative person descriptions were assigned to group A, six positive and six negative descriptions were assigned to group B (see Table 1).

4.1.3. *Procedure and data analysis*

For Experiment 2 the same experimental and analytical procedures were used as for Experiment 1.

4.2. *Results*

4.2.1. *Group assignments, frequency estimation, and evaluative trait rating*

The overall MANOVA revealed a main effect for Group (Wilk's $\Lambda = 0.73$, $F(3, 27) = 3.37$, $p = .033$). Similar to Experiment 1, the univariate follow-up analyses indicated that participants assigned more negative descriptions to the minority than to the majority ($F(1, 29) = 9.53$, $p = .012$, $\eta_p^2 = 0.25$; see Fig. 7 Top), estimated the frequency of negative traits to be higher in the minority than in the majority ($F(1, 29) = 6.26$, $p = .018$, $\eta_p^2 = 0.18$; see Fig. 7 Middle), and evaluated the minority less favorable than the majority ($F(1, 29) = 6.71$, $p = .018$, $\eta_p^2 = 0.19$; see Fig. 7 Bottom). Thus, an IC was present in all three measures. No main effect for Time point was observed (Wilk's $\Lambda = 0.92$, $F(3, 27) = 0.82$, $p = .495$). As in Experiment 1, the interaction between Group and Time point was again not significant (Wilk's $\Lambda = 0.97$, $F(3, 27) = 0.28$, $p = .840$), indicating that the IC was stable over time.

4.2.2. Memory

As in Experiment 1, arcsine-transformed unbiased hit rates were entered in a repeated measure ANOVA with the factors Group (majority vs. minority), Valence (positive vs. negative), and Time point (immediate vs. delayed). There was a significant main effect for Group ($F(1, 29) = 332.46$, $p < .001$, $\eta_p^2 = 0.92$). Source memory performance was better for the majority than for the minority (Fig. 5). This effect was modulated by Valence ($F(1, 29) = 8.44$, $p = .007$, $\eta_p^2 = 0.23$). No other effects were observed (all $Fs < 1$, $p > .413$). The Group × Valence interaction was followed up by t-tests. The follow-up *t*-test for the majority revealed a significant difference ($t(29) = 2.40$, $p = .046$, two-tailed, Cohen's $d = -0.44$) between positive and negative descriptions, with positive descriptions being better recalled than negative descriptions. As in Experiment 1, no such difference was observed for the minority ($t(29) = -1.11$, $p = .275$, two-tailed, Cohen's $d = -0.20$).

The results for Pr and Br can be seen in Fig. 6. As in Experiment 1, no effects were observed for Pr (all $Fs < 1$, $p > .649$). In contrast, the analysis of Br revealed a significant effect for Valence ($F(1, 27) = 12.65$, $p = .001$, $\eta_p^2 = 0.30$), indicating that participants were more likely to assign positive descriptions to the majority than negative descriptions. No other effects were significant (all $Fs < 1$, $p > .622$).

4.2.3. Measures of illusory correlations

Using the same procedure as in Experiment 1, correlation coefficients were calculated for each dependent variable and each time point separately. The overall MANOVA was significant (Wilk's $\Lambda = 0.57$, $F(7,23) = 2.47$, $p = .048$). As in Experiment 1, an IC was observed for all tasks but the sentence valence rating (encoded correlation: $p = .348$; Table 3). Furthermore, the encoded correlation did neither correlate with any single IC coefficient (all *p*-values $> .174$) nor with the average of all single IC coefficients ($r = -0.14$, $p = .447$).

4.2.4. Comparison between Experiment 1 and 2

In order to put the results into perspective, we also compared the results from Experiment 1 and 2 statistically. We found no differences between Experiment 1 and Experiment 2 in the bias against the minority (all $Fs < 1.23$, $p > .310$), memory task (all $Fs < 2.18$, $p > .145$), and IC coefficients (all $Fs < 1$, $p > .635$).

Next, we collapsed the samples of both experiments in order to calculate two linear regressions that assess the relative contribution of source memory performance on the ICs, separately for the frequency estimation task and the evaluative trait rating. We used the arcsine-transformed unbiased hit rates as predictors and IC as criterion. All measures were collapsed across time points. The memory performance for each of the four category combinations significantly predicted the IC in both tasks (see Table 4). As indicated by the sign of the regression coefficients, memory for positive descriptions of the majority and negative descriptions of the minority increased the IC, whereas memory for negative descriptions of the majority and positive descriptions of the minority decreased it.

Furthermore, the model was not improved, when experiment was included as categorical predictor (frequency estimation: $\Delta R^2 < 0.01$, $F(1, 52) = 0.06$, $p = .809$; evaluative trait rating: $\Delta R^2 < 0.01$, $F(1, 52) = 0.04$, $p = .845$) or as moderator (frequency estimation: $\Delta R^2 = 0.07$, $F(4, 48) = 1.55$, $p = .203$; evaluative trait rating: $\Delta R^2 = 0.02$, $F(4, 48) = 0.37$, $p = .828$).

However, as indicated by the bivariate correlations (Table 5) between the variables, memory performance for positive descriptions of the majority and for negative descriptions of the minority act as suppressor variables, i.e. they do not predict the extent of illusory correlation themselves, but contribute to the prediction by removing criterion-irrelevant variance from the other predictors (Pandey & Elliott, 2010). This might be the case due to partial structural dependence between the variables. In other words, the variance that is not shared between the variables predicts IC. Nevertheless, the results indicate that memory for both, positive and negative descriptions, plays a key role in IC.

Table 4

Standardized regression coefficients (β) of the arcsine-transformed unbiased hit rates in the prediction of the illusory correlation in the frequency estimation task and the evaluative trait rating.

	Majority		Minority	
	Positive	Negative	Positive	Negative
Frequency Estimation	0.36^{**}	-0.56^{***}	-0.39^{**}	0.28^{*}
Evaluative Trait Rating	0.53^{***}	-0.53^{***}	-0.34^{**}	0.37^{**}

Frequency estimation: $R^2 = 0.43$, evaluative trait rating: $R^2 = 0.41$.
Significantly different from zero:

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5

Correlations between unbiased hit rates and IC. Data were collapsed across the experiments.

	Frequency Estimation	Evaluative Trait Rating	Majority Positive	Majority Negative	Minority Positive	Minority Negative
Majority Positive	−0.00	0.21	–			
Majority Negative	−0.47***	−0.32*	0.35**	–		
Minority Positive	−0.43***	−0.27*	0.48***	0.44***	–	
Minority Negative	−0.01	0.11	0.08	0.48***	0.13	–

Significantly different from zero:

* $p < .05$.

** $p < .01$.

*** $p < .001$.

4.3. Discussion

Experiment 2 tested whether an IC could be observed even in the equated frequency condition. In essence, all the results from Experiment 1 could be replicated in Experiment 2.

To sum up, irrespective of the experimental condition (skewed vs. equal distribution) an IC was found. Thus, participants associated the majority with positive traits and the minority with negative traits. The IC was not affected by the delay.

5. General discussion

The main goal of the present study was to compare the relative merits of two accounts for ICs, the Shared Distinctiveness Account (SDA) and the Information Loss Account (ILA). For this purpose, we: (1) used computer simulations based on the BIAS model to derive exact predictions, whether an IC would be present in an equal frequency condition under the assumption of pure information loss, (2) tested whether the predictions derived from the simulations correspond to the behavioral data, and (3) explored whether observed illusory correlations were stable over time shortly after initial encoding.

5.1. Information loss versus shared distinctiveness

The computer simulations showed that the ILA would predict the presence of an IC in the skewed frequency condition, but not in the equal frequency condition. The main result of the two behavioral experiments is that the same extent of ICs can be observed irrespective of the frequency ratio in all three common measures of ICs (frequency estimation, group assignment, and evaluative trait rating). The latter results are clearly incompatible with the notion of the ILA that ICs can be explained by the nature of skewed frequency distributions and pure information loss. The ILA can only account for the current findings when additional assumptions are made, such as that the processing varies depending on the content or context. Fiedler (2000) offered a computational implementation of selective forgetting and weighting of information in the BIAS model. However, it would necessitate the introduction of psychological constructs like distinctiveness into the ILA that the ILA originally sought to replace. The data from the group assignment task provide some support for the ILA, insofar as source memory is better for the majority than for the minority and best for positive descriptions of the majority. Due to the preponderance of the majority and of positive descriptions, the formation of inter-item associations, which facilitate memory retrieval, are more likely for the majority than the minority (Fiedler et al., 1993).

The results of the frequency estimates, group assignments, and evaluative trait rating are consistent with the SDA, but in the group assignment task no heightened retrieval accuracy for negative descriptions of the minority was found. Based on the SDA, we expected memory to be better for the minority than the majority and to be best for negative descriptions of the minority. Instead we found the reversed pattern. Memory was better for the majority than the minority and positive descriptions of the majority were remembered best. Our findings do, therefore, not support the SDA.

The latter finding is partially consistent with Fiedler et al. (1993), who reported that memory is better for positive items compared to negative items. However, our results are inconsistent with the findings from memory studies on ICs that used multinomial processing tree models, which report that item memory is better for negative than for positive items, whereas source memory is equal for all sources (Bulli & Primi, 2006; Klauer & Meiser, 2000; Meiser, 2003).

Most IC studies only control for the desirability of the behavior descriptions (e.g. Van Rooy et al., 2013). We used trait descriptions instead of behavior descriptions. This allowed us to exert more control for factors known to affect memorability like arousal, imageability, or word length. However, the word frequency was higher for positive traits than for negative traits in our study reflecting an actual linguistic difference between positive and negative words (e.g. Zajonc, 1968). Nevertheless, the difference in word frequency between positive and negative trait adjectives cannot account for the present results, because across both experiments source memory was better for the majority than the minority for both, positive ($t(57) = 17.04$, $p < .001$, $d = 2.24$) and negative items ($t(57) = 10.81$, $p < .001$, $d = 1.42$). And last, but not least, at the immediate test negative items tended to be remembered better than positive items for the minority ($t(57) = -1.70$, $p = .095$, $d = -0.22$; see also Fig. 5).

The latter finding seems to provide some tentative support for the SDA hypothesis. However, given the small effect size, a sample of roughly 130 participants would be necessary to test this effect with sufficient power. Furthermore, as indicated by the regression analysis, only memory performance for minority members with positive descriptions and majority members with negative descriptions is a reliable predictor of IC. Positive descriptions of the majority or negative descriptions of the minority contribute to the prediction of IC only by removing criterion-irrelevant variance. Thus, even though a sufficiently powered design might provide evidence for heightened memory for shared distinctive items, this memory advantage alone might not be a strong determinant of the IC. Thus, this additional analysis also did not provide support for the SDA.

We also analyzed our data by calculating the traditional measures of items discrimination Pr and response bias Br (Snodgrass & Corwin, 1988). As major finding, the analysis indicated that positive items were more likely attributed to the majority than to the minority. This effect cannot be accounted for by differences in frequency between positive and negative items, because the response bias effect was similar in the skewed and equated frequency condition. However, this finding is in line with a study by Alves et al. (2015) on the effect of valence on recognition memory. Similar to our study, Alves and colleagues reported that there were differences in response bias, but not in item discrimination for positive and negative items. They attributed this effect to the fact that positive items are more similar to each other than negative items (Alves et al., 2015, 2017b). Thus, future studies that test the IC with trait descriptions should also control for the influence of similarity. Another, yet to be tested possibility is that the finding reflects the learning of the ratio between positive and negative items.

At first sight, the results from the traditional memory measures Pr and Br seem to be at odds with the results from the unbiased hit rates. The traditional measures revealed differences only in response bias, whereas the unbiased hit rates reveal superior memory for the majority as compared to the minority. However, these two types of measures quantify different aspects of memory. Although Pr (or d’) scores are corrected for response bias by subtracting false alarm rate (Snodgrass & Corwin, 1988), these measures assume that detection is equal for both the majority and the minority. The majority is treated as target and the minority as distractor. Genuine memory differences (i.e. differences in discrimination ability) between majority and minority would be ascribed to response bias. In the current study, we used the unbiased hit rates, because we were interested in mnemonic differences between the majority and the minority. Such differences would be masked by the traditional approaches. Unbiased hit rates allow the assessment of such memory differences and have been successfully used in previous source memory studies (Bell et al., 2012; Suzuki & Suga, 2010).

If the observed memory advantage for the majority in the unbiased hit rates was solely due to a response bias in favor for the majority, then all Br scores should be significantly larger than 0.50. Br $>$ 0.50 indicates a response bias that favors the majority over the minority. A one-sample *t*-test of the Br scores across both experiments revealed that only the Br scores for positive items were significantly different from 0.50 (immediate: $t(57) = 8.08$, $p < .001$, Cohen’s $d = 1.06$; delayed: $t(57) = 6.23$, $p < .001$, Cohen’s $d = 0.82$). The Br scores for negative items did not differ from 0.50 (immediate: $t(57) = 1.03$, $p = .306$, Cohen’s $d = 0.14$; delayed: $t(57) = 1.30$, $p = .198$, Cohen’s $d = 0.17$). These results indicate that the differences in the unbiased hit rates cannot be attributed solely to a response bias in favor for the majority, but might instead primarily reflect genuine memory differences.

A more general limitation of our experiments is that participants were informed beforehand that there will be a majority group and a minority group. Indeed, ICs based on expectancies or self-relevance have been reported even when frequencies for one dimension have been equated (Spears, Eiser, & van der Pligt, 1987; Spears, van der Pligt, & Eiser, 1986). The instructions were designed to closely follow those of Hamilton and Gifford (1976) which explicitly state the presence of a majority and a minority. This instruction might already have activated a pre-experimental association between belonging to a majority and positive traits and belonging to a minority and negative traits (McGarty et al., 1993). However, the very same instructions lead to a reversal of ICs, when negative behavior is more prevalent indicating that participants respond according to the displayed information (Hamilton & Gifford, 1976, Exp. 2). Therefore, it seems unlikely that pre-existing associations alone are responsible for the observed effect.

Future studies should use paradigms specifically designed to critically assess the validity of the SDA. The inclusion of foil items and testing the same number of items from the majority, minority and foils should make the IC paradigm less susceptible to guessing strategies and offers more possibilities for assessing memory performance. Indeed, we found heightened source memory for negative items of the minority if guessing strategy was taken into account in a recent study (Weigl, Mecklinger, & Rosburg, 2016). In addition to such improvements in the experimental design, the use of time-sensitive methods like event-related potentials or eye tracking seem suitable to shed more light on the cognitive processes at encoding that contribute to the development of ICs.

5.2. *Temporal characteristics of the illusory correlation*

No decline of ICs across time was found. In previous studies (Chapman, 1967; Lilli & Rehm, 1983) ICs wore off when participants were tested on another sequence of stimuli immediately after the IC was assessed (e.g. learning 48 stimuli twice), most likely as a result of increased transparency of the judgment situation (Lilli & Rehm, 1983, 1984). Consistent with this idea, the IC was not reduced when participants were engaged in extended learning (e.g. learning 96 stimuli in a row; Lilli & Rehm, 1983) and no IC was found, when participants were given only a summary table about the groups (Hamilton et al., 1985). In our case, participants did not run through the same experimental procedure again and could, therefore, not profit from their knowledge of the task, but had to rely on the initially encoded information.

The absence of a decline in the extent of the IC might imply that the IC helps forming an expectation, which is later used to guide behavior and attention giving rise to the expectancy-based IC (Hamilton, 1981; see also Garcia-Marques & Hamilton, 1996). Even though one might have expected to see more decay in the memory judgments and an increase in the IC itself on the basis of the ILA, all measures proved to be quite stable over the 40 min delay. This pattern of results is more consistent with the SDA. A study by Hunt (2009) indicated that the effects of distinctiveness on memory persist even after a retention interval as long as 48 h.

However, there are several limitations regarding the interpretation of the delay manipulation that should be discussed. First of all, our delay was relative short. Even though some memory decay might be expected after a short time period (e.g. Rubin & Wenzel, 1996), a longer retention interval might allow a more conclusive assessment of the stability of ICs. Second, the sample size was rather low. Thus, the absence of an effect might simply imply that the statistical power was insufficient. Third, and more critically, we measured the impact of a delay in a within subject design, i.e. subjects might have been influenced by their initial choices and ratings (e.g. via remembering or response priming). Therefore, the second assessment might not be independent from the first assessment. Furthermore, the first test offered an opportunity to further consolidate memory. Indeed, studies on the testing effect imply that repeated tests on the same material might even lead to an improved performance (e.g. Karpicke & Roediger, 2008; Rosburg, Johansson, Weigl, & Mecklinger, 2015). Future studies should also investigate the impact of delays in a between-subject design and use longer retention intervals.

5.3. Implications for other illusory correlation accounts

The current findings also inform about the validity of other proposed accounts for ICs. The mere exposure effect which is a preference for more often encountered stimuli over less often encountered stimuli (Zajonc, 1968; see also Hamilton, 1981) might offer an alternative explanation why an IC can be observed in both frequency ratio conditions. In both cases participants see more items referring to the majority than to the minority and therefore they might judge the majority group more favorably due to higher familiarity. Even though mere exposure can explain the data in our study, the mere exposure effect cannot explain the standard finding that ICs are reversed when negative items are more frequent than positive items (Hamilton, 1981, Hamilton & Gifford, 1976 Exp. 2; see also Mullen & Johnson, 1990, for a review).

The current results cannot be accounted by assuming a pseudocontingency effect (e.g. Fiedler, et al., 2009; Fiedler, Kutzner, & Vogel, 2013). Pseudocontingencies arise when a covariation judgment has to be made for two dimensions that have skewed frequency distributions as base rate (e.g. both dimensions have a ration 3:1). In contrast to ICs, joint observations of both dimensions are not necessary. Participants in a pseudocontingency paradigm use the information of the base rate to infer a correlation between two dimensions. Although the results of Experiment 1 can be interpreted as a pseudocontingency, the results of Experiment 2 are clearly incompatible with this view, because the base rate for the dimension valence is 50:50 and no pseudocontingency should arise.

Our data is also inconsistent with the accentuation approach (McGarty et al., 1993; for a related account see Sherman et al., 2009). According to the accentuation approach, participants search for meaning in the material presented in an IC experiment. The most sensible hypothesis in this scenario would be that one group is better than the other (McGarty et al., 1993). Positive information about the majority and negative information about the minority provide evidence in favor for the majority. Negative information about the majority and positive information about the minority would support the opposite conclusion. In the skewed frequency condition, there are 24 items in favor for the majority and only 12 items in favor for the minority (see diagonals in Table 1). Participants accentuate this perceived difference between the groups, i.e. the information on the diagonal favoring the majority is emphasized. In the equated frequency condition, however, the number of stimuli on the diagonals is 18 in both cases and, therefore, participants should not have a clear preference for one specific group as suggested by the accentuation approach.

This study has also implications for accounts that refute the erroneous or biased character of ICs. For example, Smith (1991) postulated that subjects in an IC paradigm rely on the absolute and not on the relative frequency in their judgment. In the original experiment of Hamilton and Gifford (1976), there is a surplus of 10 desirable behaviors for group A (18 desirable behaviors minus 8 undesirable behaviors). For group B, however, there is only a surplus of 5 desirable behaviors (9 desirable behaviors minus 4 undesirable behaviors). In this view, it is perfectly rational to rate the majority more favorably than the minority. However, in the design of Experiment 2, the surplus would be zero for both groups (12 positive traits – 12 negative traits for the majority or 6 positive traits – 6 negative traits for the minority; see Table 1). Nevertheless, an IC was observed in this condition.

A variant of this argument would be that participants might not pay attention to the complete contingency table when evaluating the groups, but instead restrict themselves to consider only the positive instances (Fiedler, 1985). Indeed, if one is explicitly asked about the number of positive instances in two classes, it is not at all erroneous to ignore the negative instances. From this perspective, both Experiment 1 and 2 would provide evidence for the positivity of the majority (16 vs. 8 in Exp. 1 and 12 vs. 6 in Exp. 2; see Table 1). This might explain why participants evaluated the majority more favorable than the minority in the evaluative trait rating, because the scale is largely composed of positive traits. However, the frequency estimation task required the participant to explicitly estimate the proportion of negative traits in both groups. In this case, participants should evaluate the majority less favorable than the minority, because there are also more negative traits in the majority than in the minority.

Our results from the source memory task are consistent with Rothbart's (1981) availability account of ICs. Availability is determined not only by distinctiveness, but also by the frequency of occurrence (Tversky & Kahneman, 1973). But in contrast to Hamilton and Gifford (1976) who focus on distinctiveness, Rothbart (1981) argues that the most frequent category combination should also be the most available category combination. In the typical IC experiment, the positive items of the majority are the most frequent stimuli. Therefore, these items should be most available in memory. Consistent with this idea, we found that source memory was better for the majority than the minority and best for positive traits of the majority. Our results are also in line with Rothbart, Fulero, Jensen, Howard, and Birrell (1978) who not only found that extreme items are more available than less extreme items (Exp. 2 & 3), but also that positive items were remembered better than negative items, if positive items are more frequent than negative items (Exp. 1). However, Rothbart's account is somewhat difficult to reconcile with the equated frequency condition in which only superior source memory for the majority, but (due to the equated frequencies) no difference between positive and negative traits would have been expected.

Finally, Murphy et al. (2011) proposed that associative learning models like the Rescorla-Wagner model (Rescorla & Wagner, 1972) can explain the development of ICs. They reasoned that the illusory correlation was only a transitory phenomenon in the acquisition stage and that the contingency judgment would be quite accurate after sufficient learning, i.e. that the IC disappears after extended learning. According to the associative learning account the IC is transitory, because learning reaches the asymptote faster for the majority than for the minority. The results of Murphy et al. (2011; see also Spiers, Love, Le Pelley, Gibb, & Murphy, 2016) support the predictions. Due to the additional distinctiveness of negative items (e.g. Alves et al., 2015, 2016, 2017b), it seems reasonable to assume that learning differs between positive and negative items. In this case, the associative learning account by Murphy et al. (2011) would also predict an IC in both, the skewed and the equated frequency condition. Critically, the illusory correlation was maximal between 36 and 54 trials in the experiments by Murphy et al. (2011). Most studies on the IC, including our two experiments, use between 36 and 48 trials (see Mullen & Johnson, 1990, for a review). It might be the case that our studies measured the pre-asymptotical state and that no illusory correlation would have been observed with more number of trials. Since we did not vary the trial number, our study does not allow conclusions about the validity of this alternative account. Future studies could address whether the illusory correlation disappears with higher trial numbers. Comparing the extent of IC between the skewed and the equated frequency condition at the asymptotic stage might provide a critical test for the validity of the associative learning account.

6. Conclusion

This study demonstrated the presence of an IC in conditions with equated frequencies which – contrary to the predictions of the information loss account – was similar in size and direction to the IC observed in the skewed frequency condition. Our results indicate that the mere frequency ratio of positive to negative information is not the only psychologically active mechanism in the distinctiveness-based IC. Rather, the asymmetry in the processing of positive and negative information (Alves, Koch, & Unkelbach, 2017b; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001) might play an important role in the formation of ICs and stereotypes. However, a pure distinctiveness approach was not supported either as rare items (negative items in the minority) were not better remembered than more frequent items. We, therefore, conclude that both accounts fall short of explaining ICs, without adding additional assumptions. The associative learning approach by Murphy et al. (2011) might offer the best explanation for the observed bias against the minority, whereas our memory data best fit best to Rothbart's (1981) availability approach. It seems that multiple factors contribute to the formation of ICs. Indeed, recent theorizing attempts to bridge the gap between the different explanatory concepts and seek to clarify their relative contribution to IC (e.g. van Rooy et al., 2013). Future studies should aim at further elucidating the role of memory and learning in the distinctiveness-based IC. On this foundation, new and more integrative models of human covariation assessment can be developed.

Acknowledgment

This research was carried out as partial fulfillment of a doctoral dissertation by the first author. A preliminary version of this study was presented at the 14th European Congress of Psychology in 2015. We would like to thank Patrizia Schneider, Julia Bock, Nicole Recklies, and Nadja Fürst for their assistance in the data collection and Prof. Dr. Dirk Wentura for his helpful comments on the data analysis and manuscript. We greatly appreciated the helpful comments of Robin Murphy and a second, anonymous reviewer.

Conflict of interest

None.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.concog.2018.06.002.

References

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *Journal of Experimental Psychology: General, 108*(4), 441–485. http://dx.doi.org/10.1037/0096-3445.108.4.441.

Alves, H., Koch, A., & Unkelbach, C. (2016). My friends are all alike—The relation between liking and perceived similarity in person perception. *Journal of Experimental Social Psychology, 62*(Supplement C), 103–117. http://dx.doi.org/10.1016/j.jesp.2015.10.011.

Alves, H., Koch, A., & Unkelbach, C. (2017a). The "common good" phenomenon: Why similarities are positive and differences are negative. *Journal of Experimental Psychology. General, 146*(4), 512–528. http://dx.doi.org/10.1037/xge0000276.

Alves, H., Koch, A., & Unkelbach, C. (2017b). Why Good Is More Alike Than Bad: Processing Implications. *Trends in Cognitive Sciences, 21*(2), 69–79. http://dx.doi.org/10.1016/j.tics.2016.12.006.

Alves, H., Unkelbach, C., Burghardt, J., Koch, A. S., Krüger, T., & Becker, V. D. (2015). A density explanation of valence asymmetries in recognition memory. *Memory & Cognition, 43*(6), 896–909. http://dx.doi.org/10.3758/s13421-015-0515-5.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370. http://dx.doi.org/10.1037/1089-2680.5.4.323.

Bell, R., Buchner, A., Erdfelder, E., Giang, T., Schain, C., & Riether, N. (2012). How specific is source memory for faces of cheaters? Evidence for categorical emotional tagging. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 38*(2), 457–472. http://dx.doi.org/10.1037/a0026017.

Bless, H., Fiedler, K., & Strack, F. (2004). *Social cognition: How individuals construct social reality.* Hove, East Sussex, UK ; New York: Psychology Press.

Bulli, F., & Primi, C. (2006). Illusory correlation and cognitive processes: A multinomial model of source-monitoring. *Review of Psychology, 13*(2), 95–102.

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 151–155. http://dx.doi.org/10.1016/S0022-5371(67)80066-5.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72*(3), 193–204. http://dx.doi.org/10.1037/h0024670.

Ebbinghaus, H. (1966). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.* Amsterdam: Bonset (Original work published 1885).

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 224–240. http://dx.doi.org/10.1037/0278-7393.21.1.224.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. http://dx.doi.org/10.3758/BRM.41.4.1149.

Feldman, J. M., Camburn, A., & Gatti, G. M. (1986). Shared distinctiveness as a source of illusory correlation in performance appraisal. *Organizational Behavior and Human Decision Processes, 37*(1), 34–59. http://dx.doi.org/10.1016/0749-5978(86)90043-9.

Fiedler, K. (1985). *Kognitive Strukturierung der sozialen Umwelt: Untersuchungen zur Wahrnehmung kontingenter Ereignisse.* Göttingen: Hogrefe.

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*(1), 24–36. http://dx.doi.org/10.1037/0022-3514.60.1.24.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*(1), 193–214. http://dx.doi.org/10.1037/0033-295X.103.1.193.

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology, 4*(1), 25–58. http://dx.doi.org/10.1037/1089-2680.4.1.25.

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology, 66*(4), 633–645. http://dx.doi.org/10.1037/0022-3514.66.4.633.

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review, 116*(1), 187–206. http://dx.doi.org/10.1037/a0014480.

Fiedler, K., Kutzner, F., & Vogel, T. (2013). Pseudocontingencies: Logically Unwarranted but Smart Inferences. *Current Directions in Psychological Science, 22*(4), 324–329. http://dx.doi.org/10.1177/0963721413480171.

Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory Correlations and Memory Performance. *Journal of Experimental Social Psychology, 29*(2), 111–136. http://dx.doi.org/10.1006/jesp.1993.1006.

Fiedler, K., & Walther, E. (2004). *Stereotyping as inductive hypothesis testing.* Hove (UK); New York: Psychology Press.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*(6), 889–906. http://dx.doi.org/10.1037/0022-3514.38.6.889.

Garcia-Marques, L., & Hamilton, D. L. (1996). Resolving the apparent discrepancy between the incongruency effect and the expectancy-based illusory correlation effect: The TRAP model. *Journal of Personality and Social Psychology, 71*(5), 845–860. http://dx.doi.org/10.1037/0022-3514.71.5.845.

Hamilton, D. L. (1981). Illusory Correlation as a Basis for Stereotyping. In D. L. Hamilton (Ed.). *Cognitive processes in stereotyping and intergroup behavior* (pp. 115–144). Hillsdale, N.J.: L. Erlbaum Associates.

Hamilton, D. L., Dugan, P. M., & Trolier, T. K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 48*(1), 5–17. http://dx.doi.org/10.1037/0022-3514.48.1.5.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*(4), 392–407.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology, 39*(5), 832–845. http://dx.doi.org/10.1037/0022-3514.39.5.832.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2*(1), 105–112. http://dx.doi.org/10.3758/BF03214414.

Hunt, R. R. (2009). Does salience facilitate longer-term retention? *Memory, 17*(1), 49–53. http://dx.doi.org/10.1080/09658210802524257.

Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(1), 81–87. http://dx.doi.org/10.1037/0278-7393.8.1.81.

Kanouse, D. E. (1984). Explaining negativity biases in evaluation and choice behavior: Theory and research. In T. C. Kinnear (Vol. Ed.), *Advances in consumer research: Vol. 11*, (pp. 703–708). Provo, UT: Association for Consumer Research.

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science, 319*(5865), 966–968. http://dx.doi.org/10.1126/science.1152408.

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin, 26*(9), 1074–1093. http://dx.doi.org/10.1177/01461672002611005.

Koch, A., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 42*(8), 1171–1192. http://dx.doi.org/10.1037/xlm0000243.

Kuhbandner, C., & Pekrun, R. (2013). Joint effects of emotion and color on memory. *Emotion, 13*(3), 375–379. http://dx.doi.org/10.1037/a0031821.

Lilli, W., & Rehm, J. (1983). Theoretische und empirische Untersuchungen zum Phänomen der “illusorischen Korrelation” (illusory correlation). I. Ableitung von Randbedingungen für das Auftreten von Effekten der illusorischen Korrelation aus dem Konzept der Verfügbarkeits-(availability-)Heuristik. *Zeitschrift für Sozialpsychologie, 14*(3), 251–261.

Lilli, W., & Rehm, J. (1984). Theoretische und empirische Untersuchungen zum Phänomen der Zusammenhangstäuschung. II. Entwicklung eines Modells zum quantitativen Urteil und Diskussion seiner Implikationen für die soziale Urteilsbildung. *Zeitschrift Für Sozialpsychologie, 15*, 60–73.

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*(4), 391–410. http://dx.doi.org/10.1002/ejsp.2420230406.

Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition, 21*(2), 121–156. http://dx.doi.org/10.1521/soco.21.2.121.21318.

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology, 29*(1), 11–28. http://dx.doi.org/10.1111/j.2044-8309.1990.tb00883.x.

Murphy, R. A., Schmeer, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. *The Quarterly Journal of Experimental Psychology, 64*(1), 24–40. http://dx.doi.org/10.1080/17470218.2010.493615.

Pandey, S., & Elliott, W. (2010). Suppressor variables in social work research: Ways to identify in multiple regression models. *Journal of the Society for Social Work and Research, 1*(1), 28–40. http://dx.doi.org/10.5243/jsswr.2010.2.

Pryor, J. B. (1986). The influence of different encoding sets upon the formation of illusory correlations and group impressions. *Personality and Social Psychology Bulletin, 12*(2), 216–226. http://dx.doi.org/10.1177/0146167286122008.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL < https://www.R-project.org/ > .

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855–863. http://dx.doi.org/10.1016/S0022-5371(67)80149-X.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.). *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Risen, J. L., Gilovich, T., & Dunning, D. (2007). One-shot illusory correlations and stereotype formation. *Personality and Social Psychology Bulletin, 33*(11), 1492–1502. http://dx.doi.org/10.1177/0146167207305862.

Rosburg, T., Johansson, M., Weigl, M., & Mecklinger, A. (2015). How does testing affect retrieval-related processes? An event-related potential (ERP) study on the short-term effects of repeated retrieval. *Cognitive, Affective, & Behavioral Neuroscience, 15*(1), 195–210. http://dx.doi.org/10.3758/s13415-014-0310-y.

Rothbart, M. (1981). Memory processes and social beliefs. In D. L. Hamilton (Ed.). *Cognitive processes in stereotyping and intergroup behavior* (pp. 145–181). Hillsdale, N.J.: L. Erlbaum Associates.

Rothbart, M., Fulero, S., Jensen, C., Howard, J., & Birrell, P. (1978). From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology, 14*(3), 237–255. http://dx.doi.org/10.1016/0022-1031(78)90013-6.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734. http://dx.doi.org/10.1037/0033-295X.103.4.734.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*(6), 523–542. http://dx.doi.org/10.3758/BF03197149.

Schmidt, S. R. (2012). *Extraordinary memories for exceptional events.* New York, NY: Psychology Press.

Shavitt, S., Sanbonmatsu, D. M., Smittipatana, S., & Posavac, S. S. (1999). Broadening the conditions for illusory correlation formation: Implications for Judging Minority Groups. *Basic and Applied Social Psychology, 21*(4), 263–279. http://dx.doi.org/10.1207/S15324834BASP2104_1.

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R. (2009). Attentional processes in stereotype formation: A common model for category accentuation and illusory correlation. *Journal of Personality and Social Psychology, 96*(2), 305–323. http://dx.doi.org/10.1037/a0013778.

Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology, 27*(2), 107–123. http://dx.doi.org/10.1016/0022-1031(91)90017-Z.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*(1), 34–50. http://dx.doi.org/10.1037/0096-3445.117.1.34.

Spears, R., Eiser, J. R., & van der Pligt, J. (1987). Further evidence for expectation-based illusory correlations. *European Journal of Social Psychology, 17*(2), 253–258. http://dx.doi.org/10.1002/ejsp.2420170211.

Spears, R., van der Pligt, J., & Eiser, J. R. (1986). Generalizing the illusory correlation effect. *Journal of Personality and Social Psychology, 51*(6), 1127–1134. http://dx.doi.org/10.1037/0022-3514.51.6.1127.

Spiers, H. J., Love, B. C., Le Pelley, M. E., Gibb, C. E., & Murphy, R. A. (2016). Anterior temporal lobe tracks the formation of prejudice. *Journal of Cognitive Neuroscience, 29*(3), 530–544. http://dx.doi.org/10.1162/jocn_a_01056.

Stroessner, S. J., Hamilton, D. L., & Mackie, D. M. (1992). Affect and stereotyping: The effect of induced mood on distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 62*(4), 564–576. http://dx.doi.org/10.1037/0022-3514.62.4.564.

Stroessner, S. J., & Plaks, J. E. (2001). Illusory correlation and stereotype formation: Tracing the arc of research over a quarter century. In G. B. Moskowitz (Ed.). *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 247–259). Mahwah, NJ: Lawrence Erlbaum Associates.

Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition, 117*(2), 224–229. http://dx.doi.org/10.1016/j.cognition.2010.08.004.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232. http://dx.doi.org/10.1016/0010-0285(73)90033-9.

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110*(3), 536–563.

Van Rooy, D., Vanhoomissen, T., & Van Overwalle, F. (2013). Illusory correlation, group size and memory. *Journal of Experimental Social Psychology, 49*(6), 1159–1167. http://dx.doi.org/10.1016/j.jesp.2013.05.006.

von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologische Forschung, 18*(1), 299–342. http://dx.doi.org/10.1007/BF02409636.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17*(1), 3–28. http://dx.doi.org/10.1007/BF00987006.

Weigl, M., Mecklinger, A., & Rosburg, T. (2016). The role of episodic memory in illusory correlation - Evidence for the distinctiveness account [Abstract]. In J. Funke, J. Rummel, & A. Voß (Eds.). *Teap 2016: Abstracts of the 58th conference of experimental psychologists* (pp. 369–370). Lengerich: Pabst Science Publishing.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(2, Pt. 2), 1–27. http://dx.doi.org/10.1037/h0025848.