Research Article

A Robust Classic

Illusory Correlations Are Maintained Under Extended Operant Learning

Florian Kutzner,[1] Tobias Vogel,[2] Peter Freytag,[1] and Klaus Fiedler[1]

[1]University of Heidelberg, Germany, [2]University of Mannheim, Germany

Abstract. In the present research, we argue for the robustness of illusory correlations (ICs, Hamilton & Gifford, 1976) regarding two boundary conditions suggested in previous research. First, we argue that ICs are maintained under extended experience. Using simulations, we derive conflicting predictions. Whereas noise-based accounts predict ICs to be maintained (Fielder, 2000; Smith, 1991), a prominent account based on discrepancy-reducing feedback learning predicts ICs to disappear (Van Rooy et al., 2003). An experiment involving 320 observations with majority and minority members supports the claim that ICs are maintained. Second, we show that actively using the stereotype to make predictions that are met with reward and punishment does not eliminate the bias. In addition, participants' operant reactions afford a novel online measure of ICs. In sum, our findings highlight the robustness of ICs that can be explained as a result of unbiased but noisy learning.

Keywords: skewed base rates, matching to sample

It is amazing that stereotypes can form without any kernel of truth, sentiment or real group conflict. In their seminal work on frequency-based illusory correlations (ICs), Hamilton and Gifford (1976) presented participants with a series of statements describing positive or negative behaviors performed by individuals from two fictitious groups. The series of statements had three key characteristics: (a) positive behavior descriptions appeared more often than negative ones, (b) one group appeared more often than the other, and (c) the ratio of positive to negative behavior descriptions was the same for both groups, setting the correlation between groups and valence to zero. Thus, there was no apparent reason to evaluate the two groups differently. Yet, subsequent group impression ratings, frequency estimates, and memory-based assignments of the behavior descriptions to the groups reflected more positive evaluations of the majority than the minority.

The formation of illusory stereotypes differentiating majorities from minorities has since been demonstrated in numerous studies (for a review see Mullen & Johnson, 1990). ICs have been shown to generalize to environments where the frequent valence is negative resulting in more positive minority evaluations (e.g., Hamilton & Gifford, 1976, Exp. 2), to nonevaluative attributes (Sherman et al., 2009) and to more complex (Meiser & Hewstone, 2001) or more asymmetric distributions (Chun & Lee, 1999; Shavitt, Sanbonmatsu, Smittipatana, & Posavac, 1999). However, despite this seemingly robust nature of the IC phenomenon, and despite wide agreement on its important theoretical implications, prior research also suggests limitations that restrict the domain of IC effects in reality.

We are particularly concerned with two limitations that are implicit in the study designs of almost all prior IC experiments. In short, one might conclude that ICs describe stereotype formation but not stereotype maintenance since there is no evidence for ICs involving substantially more than 40 observations. Additionally, even if ICs should prove persistent in the laboratory, one might argue that in reality consequential feedback of the utilization of a faulty impression would soon eliminate the bias. Challenging both of these implicit limitations, we report new empirical evidence for ICs under extended operant-learning conditions and use computer simulations of two prominent learning models of ICs to understand why ICs are maintained and applicable to feedback-learning settings.

Illusory Stereotype Formation and Maintenance

That ICs were related only to stereotype formation, in contrast to stereotype maintenance, has already been pointed out by Mullen and Johnson (1990). In their meta-analytic review on ICs they conclude that "*... in everyday processing of information about different groups engaging in different types of behavior, the individual is likely to be exposed to many times the 40 exemplars which subjects saw in the laboratory experiment*" (Mullen & Johnson, 1990, p. 24). In fact, across all 26 reviewed studies the minority was, on average, observed just 12 times per study, giving rise to speculations ranging from reduced (Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003,

DOI: 10.1027/1618-3169/a000112

p. 540) to increased (Mullen & Johnson, 1990, p. 14) IC effects with extended observations. To our knowledge, no study has since addressed the question of whether extended experience changes the IC effect.

Finding evidence for persisting ICs would greatly expand the scope of the illusion as it applies to real-life stereotyping of minorities. Not only stereotype formation, but also stereotype maintenance regarding long-standing or historic minorities could be explained without assuming real group differences, motivated or expectancy-driven persistence (Kunda & Oleson, 1995; Snyder & Swann, 1978) or real group conflict (Sherif, 1967).

Persisting ICs are also of theoretical interest because different accounts, developed to explain the formation of ICs, diverge in their predictions for long-term maintenance. Several models share the notion that the formation of ICs can result from unbiased learning but differ in the reasons they highlight (Fiedler, 2000; Smith, 1991; Van Rooy et al., 2003). We use these models to exemplify how different assumptions about associative learning lead to different predictions for ICs under extended experience. Whereas noise results in persisting ICs (Fiedler, 2000; Smith, 1991), focusing on discrepancy-reducing feedback learning makes ICs disappear (Van Rooy et al., 2003). In sum, the first goal of the present research is to empirically explore whether illusory stereotypes persist. In doing so, we additionally explore which of two aspects of learning, noise or discrepancy-reducing feedback learning is suited better to describe ICs under extended experience.

ICs Despite Consequential Feedback

Considering maintained stereotypes, another issue becomes pressing. What happens to ICs when eventually used to make predictions (Crocker, 1981)? In fact, the act of predicting is central to one of the most prominent definitions of stereotypes "*as probabilistic predictions that distinguish the stereotyped group from others*" (McCauley & Stitt, 1978, p. 929) and under many conditions people who form attitudes might by adequately characterized as actively involved as social interactants whose correct and incorrect predictions of others' behaviors are met with benefits and costs, respectively (Denrell, 2005; Fazio, Eiser, & Shook, 2004).

In the vast majority of all previous experiments, however, IC formation is studied as an observation task, whereby the participant never uses the illusory stereotype. Even though IC effects have been demonstrated with participants actively trying to form impressions,[1] either by mere observation (Fiedler, Russer, & Gramm, 1993; Meiser, 2003) or by actively testing hypotheses (Fiedler, Walther, Freytag, & Plessner, 2002; Fiedler, Walther, & Nickel, 1999), evidence for ICs under conditions where predictions are made and met with consequences is still missing (for a notable exception see Eder, Fiedler, & Hamm-Eder, 2011). This suggests that the illusion disappears when operant elements and continuous predictions force participants to actively revise their impressions of the target groups.

Overview of Empirical Demonstrations

We believe there is reason to expect ICs even in the face of continuous predictions and consequential feedback. As mentioned above, recent accounts have emphasized that ICs can form as a byproduct of unbiased learning (Fiedler, 2000; Smith, 1991; Van Rooy et al., 2003). Importantly, the underlying conceptual frameworks, distributed memory (e.g., Hintzman, 1986), and the formation of associative links (e.g., Rescorla & Wagner, 1972), have regularly been applied to describe learning with consequential feedback, namely operant learning in humans and animals (for a recent review see Mitchell, De Houwer, & Lovibond, 2009). Thus, there seems to be no reason to restrict the domain of ICs to more or less passive observation tasks without consequential utilization of the group stereotype.

To substantiate that ICs are persistent and applicable to both passive observation and learning with predictions and consequential feedback, we first report two computer simulations of models that have been shown to explain the formation of ICs and newly explore their predictions for extended experience. As we will show, both – maintenance and disappearance of ICs – can be predicted focusing on different aspects of an unbiased learning process. Then, we empirically test the robustness of ICs with respect to extended experience and active usage. In a pre-test with the standard number of observations we establish a baseline for the IC effect. In the main study, participants evaluate majority and minority after having made 320 observations, nearly 10 times the amount of information usually provided. Additionally, for half of the participants we introduce a new operant-learning variant of the IC paradigm. This includes predicting the valence of behaviors expected from members of the majority and the minority group and a monetary incentive for correct predictions.

Simulation 1: Noise and Extended Experience

Smith (1991) relied on a distributed-memory framework (Hintzman, 1986) to explain the formation of ICs.[2] A central assumption of this account is that noise reduces the reliability of the entire memory, which is the basis for judgments. Every aspect of an observation made, be it the specific

[1] We are aware of the conflicting point of view that ICs might vanish under conditions that foster impression formation (Hamilton & Sherman, 1989; McConnell, Sherman, & Hamilton, 1994; Pryor, 1986; Sanbonmatsu, Sherman, & Hamilton, 1987). We refer the reader to the General Discussion for an attempt to integrate conflicting evidence.

[2] For the purpose of illustration we only present a replication of the Smith (1991) model. Identical conclusions follow from another distributed-memory model, the BIAS model (Fiedler, 2000).

behavior, valence or group membership, has a certain chance of being partially or completely lost. Similar to the concept of reliability in measurement, any systematic trend is therefore more accurately extracted when based on a larger number of observations. This increase in reliability is sufficient to explain the formation of ICs: There are comparatively more observations for the majority than for the minority, creating a more reliable memory for the majority. Therefore, the memory for the majority will reflect the prevalent valence to a higher degree than the memory for the minority.

At first sight, one might suspect that, given enough observations, this noise-based account predicts disappearing ICs. If the number of observations was large enough, perfect learning for majority and minority should result in equivalent evaluations for both groups, in accordance with the prevalent valence in the stimulus series. However, the assumption that noise affects every aspect of the observations made, including group membership, results in the prediction of persisting ICs.

As illustration, the positivity of the minority is determined by assessing to what degree the memory for minority members on average reflects positive valence. To obtain this average minority memory, all individual memory traces are weighted by their similarity with the prototype for the minority and then summed up (see Appendix for a computational example). As a consequence, traces similar to the minority (usually traces actually stemming from observations with the minority) contribute more to the average minority memory than atypical ones. When memory is perfect, the average minority memory only contains minority memory traces since minority traces are "positively" similar and the similarity is virtually zero for traces stemming from the majority. In contrast, when memory is noisy, the similarity for traces stemming from the majority can be different from zero. When (positively) similar, a majority trace simply acts as a minority trace and is added to the average minority memory with its original valence. When "negatively" similar, however, the trace enters the average minority memory as if everything involving this observation was just the antipode of the minority. Importantly, this includes the observed valence (see example trace in Appendix). Thus, when a positive majority memory trace is taken to imply the antipode of the minority, its valence is inverted and added to the average minority memory as a negative observation. Since accidentally creating an antipode of a majority trace with the frequent valence is more prevalent than all other such confusions, the average of minority memories will be biased toward the infrequent valence more strongly than the average of majority memories, resulting in persisting ICs.

For a test of these considerations, we expanded the simulation approach by Smith (1991). We replicated the simulation 20 times, successively increasing the number of observations from 16 to 320. The distribution we used maintained all crucial properties of ICs. One of two groups was more frequently observed than the other, either positive or negative valence dominated, and the ratio of positive to negative observations was the same for both groups (see Table 1). To demonstrate the central role of noise, we varied the chance with which individual aspects of memory were lost in four steps from 0 to 90%.

Table 1. Frequency table indicating the stimulus distribution used in the simulations and experiments

	Frequent valence	Infrequent valence	
Majority	9	3	12
Minority	3	1	4
	12	4	16

Note. One group and one valence are more frequent than the respective other at a ratio of 3:1. The contingency between group and valence is zero.

Identical to Smith, every memory trace consisted of 26 features encoding group membership (6 features), the specific group members (6 features), valence (7 features), and the specific behavior (7 features). For every simulation run, the group-membership segments and the segment for the frequent valence were created by randomly selecting feature values of -1 and 1. The segment for the opposing, infrequent valence was then computed by multiplying all of the seven valence features with -1. As dependent measure, we used the correlations between the valence-segments of the memory echo, elicited either by a majority or by a minority prompt, and the prototype for the frequent valence (for computational details see Smith, 1991, p. 121). Thus, evaluations of the groups ranged from -1, the group being characterized by the complete opposite of the frequent valence, to 1, the group being characterized by the frequent valence. Figure 1 presents the simulated group evaluations (circles) and their differences (triangles) averaged over 1000 simulation runs as a function of number of observations and different levels of noise, including 30% the value used by Smith (1991, solid symbols).

There are two main characteristics. First, noise is the driving force behind the size of the IC effect (i.e., the difference between majority and minority evaluations, triangles in Figure 1), corresponding nicely with new empirical evidence for more pronounced IC effects when working memory is impaired (Eder et al., 2011). In the absence of noise, majority and minority evaluations both reflect the prevalent valence to a degree that leaves little room for unequal evaluations. In contrast, with increasing levels of noise the evaluations for the minority become less pronounced than for the majority, creating substantial differences in evaluations, IC effects. Second and central to the present endeavor, noise causes ICs to persist. For all noise levels, the size of the IC effect changes for the first half of blocks at maximum and then stabilizes at an asymptote.

In sum, the distributed-memory model by Smith (1991) serves to illustrate how noise in otherwise unbiased information processing can cause the formation and maintenance of ICs. Under ideal conditions without noise, when memory is perfect, majority and minority evaluations are predicted to be similar. When memory is noisy, impairment in memory is less detrimental for the majority than for the minority, leading to a more complete extraction of the prevalent valence, positive or negative, for the majority. Increasing the

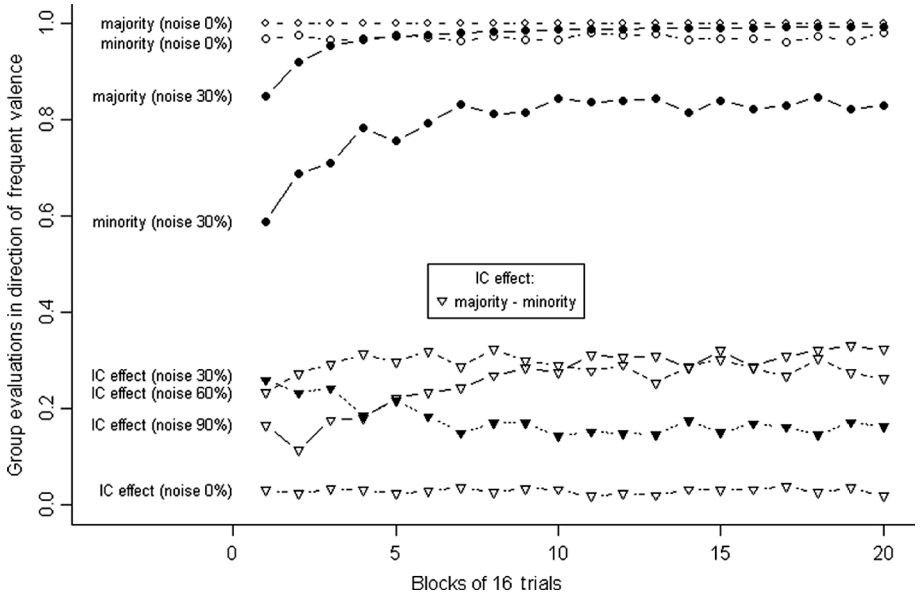

Figure 1. Simulated group evaluations, averaged across 1,000 simulated runs in a noise-based account for ICs (Smith, 1991) applied to the stimulus distribution form Table 1. Evaluations are plotted as a function of the number of observations and the amount of noise in memory.

absolute number of observations does not eliminate this advantage of the majority. Thus, the presence of noise warrants in persisting ICs.

Simulation 2: The Δ-Algorithm and Extended Experience

The conflicting prediction that ICs disappear with increasing numbers of observations can also be derived when focusing on another feature of unbiased learning, discrepancy-reducing feedback learning that converges to an asymptote. A prominent implementation of this idea is provided by Van Rooy and colleagues (2003) that relied on a recurrent-connectionist model (McClelland & Rumelhart, 1988) to explain, among other phenomena, the formation of ICs. The model learns to represent the observations made in the connections between the observed concepts, adjusting the connections' strengths. For example, frequently observing positive behavior of majority members will successively strengthen the connections between majority and positive. Central to this account, feedback learning is governed by a discrepancy-reducing Δ-algorithm (McClelland & Rumelhart, 1988, p. 166), the essence of which can be formalized as

$$learning = learning\,rate \times (observation - expectancy),$$

where *learning* stands for the change in the connections between the concepts such as positive valence. The Δ-algorithm implies that the discrepancy between what is observed and what is expected due to previous learning decreases steadily. With every observation, the discrepancy decreases by a constant proportion of its size, the learning rate.

The Δ-algorithm has two critical implications for ICs. First, the model can account for the formation of ICs. Initially, as long as the discrepancy between the observations and the model is large, changes in the strength of connections are large. Later on, as the discrepancy decreases, changes become smaller. Because the majority is observed more frequently than the minority, and because one valence is observed more frequently than the other, the connections between the majority and the frequent valence strengthen earlier than all other connections. This implies that, initially, the representation of the majority more strongly reflects the more frequent over the infrequent valence than the representation of the minority. Secondly, the Δ-algorithm implies that, after many observations with either the majority or the minority, learning for both groups approaches the same asymptote, so that ICs should disappear with sufficient observations. Because both groups are repeatedly presented with positive and negative observations, the groups' connections with positive and negative valence will approach the same asymptote.

For an illustration, we extended the simulation by Van Rooy and colleagues (2003), increasing the number of observations in 20 steps from 16 to 320 (cf. Table 1). In detail, the model consisted of nodes for the majority, the minority, positive and negative valence, and one node for every specific observation made. All nodes were bi-directionally connected with all other nodes. We ran a single internal updating cycle. Identical to VanRooy and colleagues' model, values of external activations were set to 1 and 0 and the learning rate was set to .15.[3] Group evaluations were computed from the activations of the valence

[3] In simulations, not reported here, we found the implications regarding the size of the IC effect to be robust varying the learning rate from .01 to .2.

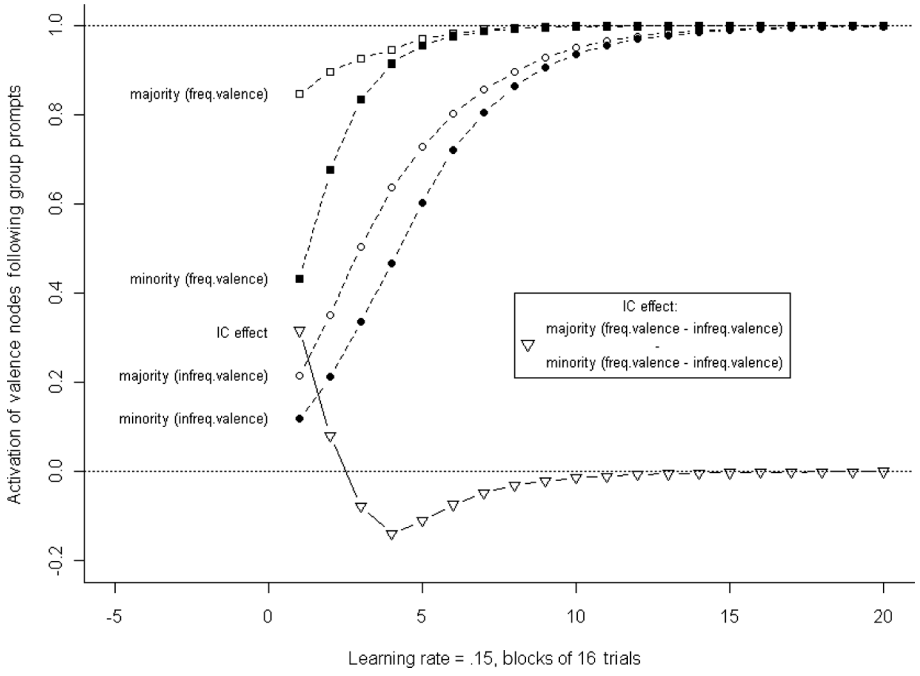

Figure 2. Simulated group evaluations, averaged across 250 simulated runs in a discrepancy-reducing feedback-learning account for ICs (Van Rooy et al., 2003) applied to the stimulus distribution from Table 1. Activation levels in the valence nodes and the resulting IC effect are plotted as a function of the group prompt (majority and minority) and the number of observations.

nodes resulting from prompting the network with the groups. Specifically, the groups' evaluations in the direction of the frequent valence were computed as the difference in activations of the frequent minus the infrequent valence node given the group prompt (for further computational details see Van Overwalle & Siebler, 2005, p. 272, ff.). We repeated the simulation 250 times for every block of 16 trials.

The results of the simulation provide a typical example of discrepancy-reducing feedback learning as computed by the Δ-algorithm. Figure 2 depicts the activation level in the valence nodes. This activation represents how strongly the network expects the frequent valence (squares) and the infrequent valence (circles) given a majority (white squares and circles) or a minority (black squares and circles) group member. As implied by the Δ-algorithm, the main determinant of expecting a valence is the number of observations with that valence. As evident from the position of the squares in Figure 2, the frequent valence is expected to a stronger degree than the infrequent valence. Analogously, both valences are expected to a stronger degree for the majority than for the minority. Finally, all expectations approach the same asymptote.

These results also provide the clear prediction that the size of the IC effect (triangles) should change as a function of the number of observations. Specifically, during early stages of the learning process the majority evaluation is more strongly in accordance with the frequent as compared to the infrequent valence than the minority evaluation, resulting in an IC effect. As observations become more numerous, this tendency slightly reverses (cf. triangles in Figure 2). This is due to the fact that the network already expects both valences to be associated with the majority to a similar, nearly maximal degree but for the minority the association with the frequent valence still enjoys an advantage. Finally, however, the IC effect is bound to vanish as all expectations reach their asymptote.

In sum, explaining ICs with distributed memory and feedback-learning models not only illustrates that it is of theoretical interest to study long-term IC effects. By analogy, it also suggests that ICs can be expected under the same wide range of conditions that have been successfully described by these learning models, including incidental and intentional learning as well as operant feedback learning.

ICs in Extended Operant Learning

Our experimental test of the robustness of ICs was similar to the standard frequency-based IC paradigms used by Hamilton and Gifford (1976). The study comprised two phases, a learning phase and an evaluation phase. In every trial of the learning phase, the label for one of two groups was presented, with one group being more frequent than the other, followed by a positive or negative behavior description. Identical to the simulations, the joint frequencies of group labels and behaviors followed the distribution depicted in Table 1 for half of the subjects. To generalize our findings beyond the association of minority with negative, for the other half, the frequently presented valence was negative. In a subsequent evaluation phase, attitude ratings regarding the groups were assessed.

To test whether ICs persist, we increased the number of observations from a maximum of 48, the longest learning sequence found in Mullen and Johnson's (1990) meta-analysis, to 320, 20 times the stimulus distribution displayed in Table 1. If ICs remain stable after such an extended number of observations, this would greatly expand the generality and the domain of application of the illusion. As the

preceding simulations show, such a persisting IC effect would be compatible with a noisy learning process but not with an asymptotic feedback-learning process.

Additionally, we wanted to test whether active involvement, making predictions that convey benefits and costs, would eliminate ICs. Therefore, we created a new operant learning task. On every trial, participants expressed their expectations about the two groups. From the first trial on, before the behaviors could be observed, participants were asked to use the group information to make predictions about the valence of the next behavior. To make predictions carry consequences, we introduced reinforcement after every response, consisting of the corrective feedback "correct" or "false" and a reward of €0.03 (∼ $0.04) for each correct response and a €0.03 punishment for each false response. Finally, only after all 320 predictions had been made, participants were once asked to explicitly evaluate the groups.

Because we measured the stereotype only once, one might argue that any effect was on the decline and that further experience would have completely eliminated it. One way to deal with this objection is to continuously measure the IC effect and to extrapolate its development. Luckily, a byproduct of the operant-learning task is such a continuous online measure. As participants can only use group membership as predictor, a difference between the groups in the ratio of predicting positive to negative valence can be interpreted as an online measure of the subjective stereotypes (Allan, 1993; McCauley & Stitt, 1978). For example, the majority is positively stereotyped if predicting positive valence is more likely for majority than for minority members. By extrapolating the resulting learning curve, we expected to substantiate our claim that extended experience would not eliminate the IC effect.

Pretest

We conducted a pretest ($N = 36$) to ensure our procedure and measures would result in a standard IC effect as typically found in studies using limited numbers of observations. Thus, we restricted the number of observations to 48 and instructed participants to form an impression of the two groups. As for the main study, we also varied whether positive or negative valence was more frequent. After the learning phase, participants indicated their overall impression of the two groups. Using scroll bars, participants rated the groups' likeability and indicated their willingness to spend time with members of each group. We formed an overall evaluation measure by averaging both ratings for the majority ($r(36) = .90$, $p < .001$) and the minority ($r(36) = .73$, $p < .001$). The rest of the procedure was identical to the main study.

A two-factorial mixed analysis of variance (ANOVA) with the valence of frequent behavior (positive vs. negative) as the between-subject factor and type of group as the within-subject factor (majority vs. minority) only revealed a marked two-way-interaction between group and valence, $F(1, 34) = 11.61$, $p < .01$, $\eta^2 = .26$ (all other F's < 1). This reflects an IC effect in that the majority ($M = 60.2$, $SD = 22.4$) was evaluated more positively than the minority ($M = 43.6$, $SD = 21.1$) if the more frequent behavior was positive, $t(1,18) = 1.87$, $p < .05$ (one tailed), whereas the minority ($M = 59.8$, $SD = 14.8$) was evaluated more positively than the majority ($M = 37.2$, $SD = 20.3$) when the less frequent behavior was positive, $t(1,16) = -3.20$, $p < .01$. The prevalent valence did not significantly influence the size of the IC effect (i.e. the absolute difference between majority and minority evaluations), $F(1, 34) < 1$. These results are yet another demonstration that the illusion is robust irrespective of processing instructions to form impressions and establish a baseline for the IC effect with our materials and procedure (Table 2).

Table 2. Evaluations (on a scale from 0 to 100) of majority and minority as a function of the type of the more frequent behavior (positive or negative)

	Frequent valence positive		Frequent valence negative	
	Majority	Minority	Majority	Minority
M	61.1	40.4	28.4	53.1
SD	21.1	22.3	20.4	23.5

Method

Participants and Design

Fifty-eight undergraduate students (32 female, 26 male) from the University of Mannheim participated in the study for compensation of €4. Subjects were randomly assigned to one of four conditions in which either positive or negative behaviors were more frequent, both at a ratio of three to one, and in which they were either required to make predictions or not. This resulted in a 2 (frequent behavior: positive vs. negative) × 2 (task: observation vs. prediction) × 2 (majority vs. minority) mixed factorial design with repeated measures on the last factor.

Procedure

Participants were informed that the study investigated social perception and that they should form an impression of two groups. We referred to the groups as the "Purple Group" and the "Orange Group". The colors of the groups were counterbalanced across participants. When the group labels appeared on the screen, they were accompanied by a corresponding hue as background of the screen and the keyboard was locked for 1 s. To move on, participants pressed one of the two response keys on the left and right side ("A" and "Ä", respectively) of a German computer keyboard and a positive or negative behavior description appeared next to the group label for 3 s. A blank screen of 0.3 s separated the observations.

Participants in the prediction condition were additionally informed that they could win money by correctly predicting the valence of the behavior performed by members from two different groups. Predictions were made on the same two keys as for the participants in the observation condition which were now labeled "positive" and "negative". The meaning of the keys remained constant over the course of the experiment but was counterbalanced between participants. After every prediction, the behavior description appeared on the screen and remained visible for 3 s together with the group label and the feedback. If predictions matched the predetermined valence of the trial, €0.03 were added to the participants compensation and the label "correct" appeared on the center of the screen. If responses did not reflect the predetermined valence of the behavioral description, €0.03 were subtracted and the label "false" appeared. The updated value of the compensation and the last change were constantly displayed in the lower right corner of the screen.

We used 128 behavior descriptions that had been pretested to have moderately positive or negative valence and described mainly other relevant behaviors (Peeters & Czapinski, 1990). For example, one negative behavior description read, "A member of the [purple/orange] group uses vast amounts of pesticides in his garden," and a positive one "A member of the [purple/orange] group can laugh about his own weaknesses." To compose the sequence of 320 trials, for every subject the computer generated 20 random sequences, each consisting of 16 trials that retained the statistical properties of the distribution in Table 1. For every participant, we randomly selected behavior descriptions from the pool. These 20 sequences were adjoined to produce the 320-trial sequence. After this learning phase, participants indicated their overall impression of the groups on two items for the majority, $r(58) = .89$, $p < .001$, and for the minority, $r(58) = .71$, $p < .001$.

Results and Discussion

End-of-Sequence Evaluations

We submitted the evaluations of the groups, which we obtained after the learning phase to a three-factorial mixed ANOVA with valence of frequent behavior (positive vs. negative) and task (observation vs. prediction) as the between-subject factors and type of group as the within-subject factor (majority vs. minority). This analysis revealed a main effect for the valence factor, $F(1, 54) = 9.12$, $p < .01$, $\eta^2 = .14$, indicating that the groups were evaluated more positively when the frequent type of behavior was positive.

Crucially, it also revealed an IC effect manifested in the interaction between valence and group, $F(1, 54) = 16.63$, $p < .001$, $\eta^2 = .24$. The majority ($M = 61.1$, $SD = 21.1$) was evaluated more positively than the minority ($M = 40.4$, $SD = 22.3$) when the more frequent behavior was positive, $t(1, 28) = 2.67$, $p < .05$, but more negatively ($M = 28.4$, $SD = 20.4$) than the minority ($M = 53.1$, $SD = 23.5$) when the less frequent behavior was positive, $t(1, 28) = -4.08$, $p < .001$. The size of the IC effect was unaffected by the task, $F(1, 54) = 1.38$, $p = .25$, valence, $F(1, 54) < 1$, or their interaction $F(1, 54) < 1$.

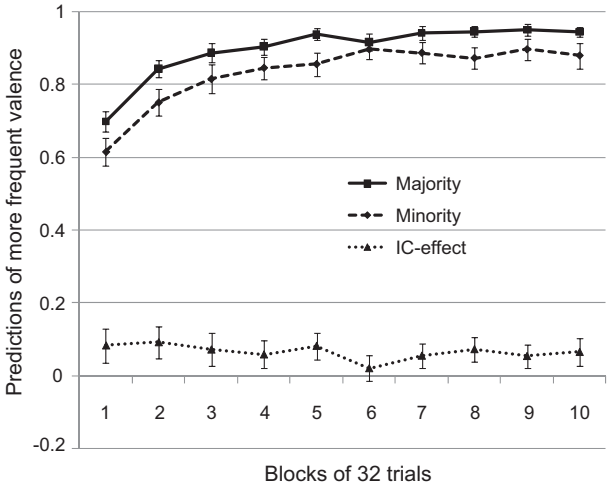

Figure 3. Mean conditional proportions (solid and dashed lines) and their difference, the IC effect (dotted line), for predicting the more frequent valence as a function of block (1–10) and the type of group (majority or minority). Error bars represent the standard error of the mean.

Online Responses

For participants who made predictions, we monitored the development of the IC effect. We analyzed the online predictions by averaging within 10 consecutive blocks of 32 trials. We computed the proportion of predictions for the more frequently presented valence, given majority and minority. These proportions were submitted to a three-factorial mixed ANOVA with valence of frequent behavior as the between-subject factor (positive vs. negative) and blocks (1–10) and type of group (majority vs. minority) as within-subject factors.

The analysis revealed a strong main effect for the within-subject factor blocks, $F(9, 27) = 15.70$, $p < .001$, $\eta^2 = .84$. Over the course of learning, the more prevalent valence was predicted more and more frequently, reaching rates of approximately .9 (see Figure 3, solid and dashed lines). It also yielded a tendency for the between-subject factor valence, $F(1, 35) = 3.88$, $p = .06$, $\eta^2 = .10$, indicating a more pronounced tendency to predict the frequent type of behavior when the frequent type was positive.

Crucially, an IC effect in the online predictions could be observed as a main effect for type of group, $F(1, 35) = 14.45$, $p < .001$, $\eta^2 = .29$. Choosing the more frequent valence was more pronounced for the frequent than for the infrequent group. No significant interaction between this online IC effect and valence emerged, $F(1, 35) = 1.39$, $p = .25$, $\eta^2 = .04$.

To analyze the course of the IC effect we computed a difference measure of the online ICs, subtracting the proportions of predictions for the frequent type of behavior given the minority from the proportions given the majority. These

differences capture whether the majority is more closely associated with the prevalent valence than the minority (see Figure 3, dotted line). We compared the size of the effect after the first block to the size after the last block. Including repeated measures for the first block and the last block revealed no evidence for a change in the size of the IC effect, $F(1, 35) < 1$. We also tested whether the IC effect followed a linear or a quadratic course. We submitted the IC measure to a repeated measures ANOVA including a linear and a quadratic contrast for the within-subject factor blocks (1–10). The results showed no evidence for either a linear or a quadratic course of learning as we did not find evidence for a linear or a quadratic component in the course of the IC effect, $F < 1$.

Finally, to assess the correspondence between the online and the end-of-sequence IC effects, we computed an overall online score by averaging the 10 successive difference scores ($\alpha = .79$). This overall online IC effect substantially correlated with the size of the ICs in the end-of-sequence evaluations, $r(37) = .50$, $p < .01$.

Deviating from standard illusory correlation paradigms, we drastically increased the number of observations to examine the persistence of ICs under extended experience. Even after 320 trials we found persisting and stable IC effects irrespective of whether participants merely observed information about the groups or actively used the stereotypes to make predictions. A preference persisted for the majority when positive behavior was predominant and for the minority when negative behavior was predominant. The IC effect ($\eta^2 = .24$) was of virtually the same size as in the standard-length pretest ($\eta^2 = .26$).

Further evidence for the robustness of ICs stems from the participants' predictions. Participants indirectly provided evaluations by continuously predicting more of the frequent valence for the majority than for the minority. This novel behavioral online measure corresponded meaningfully to the end-of-sequence evaluations and, due to its stability, gives no reason to believe that the effect was on the decline.

One might argue, however, that the null effect for the amount of experience was due to a lack of sensitivity of our online stereotype measure. Possibly participants formed an impression early on that they were reluctant to change. Therefore, it is important to note that not every effect involving the number of trials was null. To the contrary, the number of trials had the largest effect in the current study, majority and minority evaluations increasingly reflecting the frequent valence. This is evidence for the sensitivity of the online stereotype measure for participants' changing impressions and further substantiates our claim that ICs are robust to extended experience.

Finally, inspecting the resulting learning curve also helps to distinguish which of the two principles propagated by the prominent models, noise or asymptotic feedback learning, better accounts for ICs under extended experience. Strictly speaking, finding IC effects of similar size after 48 and 320 observations is not only compatible with the predictions of the noise-based learning models (e.g., with a noise of 60%, cf. Figure 1). It is also compatible with the discrepancy-reducing feedback-learning models when assuming a very slow learning rate of around .02. Under these conditions, the initial decline in the IC effect in Figure 2 is preceded by an even earlier increase resulting in an inverted-U-shaped learning course overall. However, the development of the size of the ICs over increasing numbers of observations can set the two underlying learning mechanisms apart. Whereas focusing on noise predicts a largely stable effect, asymptotic feedback learning predicts, depending on the learning rate, a decreasing or inverted-U-shaped change of the size of the IC effect. Thus, the stable nature of the IC effect over the course of learning is distinctly predicted by the noise-based account. Additionally, extrapolating the finding that the amount of learning had no effect on the size of ICs suggests that further training would not have eliminated it.

General Discussion

In the present research, illusory correlations (ICs, Hamilton & Gifford, 1976), the stereotype that majorities are better described by what is frequent than minorities, are shown to be more general and hence more widely applicable in reality than commonly assumed. New empirical evidence suggests that ICs do not fade out with extended experience and that they are robust against actively forming and using the stereotypic impression, a finding easily explained by learning models assuming random noise in information processing.

We provide empirical evidence for maintained ICs under extended experience. Still, after having observed 320 positive and negative behaviors by majority and minority members, group evaluations reflected a preference for the majority given predominantly positive behaviors, and a preference for the minority given predominantly negative behaviors. As compared to a standard length condition of 48 trials, the size of the effect seemed to remain unchanged and, continuously monitored, was largely stable. Thus, the sheer numerical predominance of one group and one valence is sufficient not only for illusory stereotypes to form but also for the same stereotypes to be maintained.

Theoretically, we show that formation and maintenance of ICs can both be explained as an emergent phenomenon of unbiased learning. Simulations based on a distributed-memory model (Hintzman, 1986) show that assuming random noise is the key to create systematically biased group evaluations in IC settings (Fiedler, 2000; Smith, 1991). These models also predict the maintenance of ICs. Due to noise, frequent majority memories are sometimes mistaken for the antipode of what the minority stands for and vice versa. Due to its numerical predominance, specific majority memories distort the average minority memory to a stronger degree than specific minority memories distort the average majority memory.

In contrast, another learning mechanism that predicts the formation of illusory stereotypes, discrepancy-reducing feedback learning, seems unsuited to explain the maintenance of ICs. Simulations of a recurrent-connectionist model that implements asymptotic feedback learning (McClelland & Rumelhart, 1988) predicted that illusory stereotypes follow a curvilinear course and disappear after many learn trials

(see also, Van Rooy et al., 2003, p. 540). This prediction was not supported as the size of the IC illusion proved largely stable. In sum, noise in the unbiased processing of social information seems to offer a parsimonious explanation of both, stereotype formation and maintenance.

We also add to the empirical evidence showing that ICs persist under conditions that facilitate online impression formation. Similar to previous research, we directly instructed all our participants to form impressions (Eder et al., 2011; Fiedler et al., 1993; Fiedler et al., 1999; Fiedler et al., 2002; Meiser, 2003). In contrast, some evidence and theorizing supports the notion that ICs cannot occur under impression formation conditions. Based on the original distinctiveness explanation (Hamilton & Gifford, 1976) and on reduced IC effects under impression formation instructions (Hamilton & Sherman, 1989; McConnell et al., 1994; Pryor, 1986; Sanbonmatsu et al., 1987), it has been suggested that the IC illusion is confined to memory-task instructions.

One way to reconcile the conflicting findings with regard to processing goals is offered by the noise-based account to ICs. Studies of person memory in general (Hamilton, Katz, & Leirer, 1980; Hastie & Park, 1986; Lichtenstein & Srull, 1987; McConnell, 2001; Srull, 1981) and of memory in IC paradigms in particular (Meiser, 2003) suggest that impression formation instructions reduce noise and enhance memory performance. However, when explained by noise in unbiased processing (Fiedler, 2000; Smith, 1991), ICs diminish as memory quality increases (Eder et al., 2011). With the appropriate caution of interpreting null findings, previous failures to obtain ICs under impression formation instructions (Hamilton & Sherman, 1989; McConnell et al., 1994; Pryor, 1986) might stem from excessively good memory. In sum, there seems to be no apparent reason to restrict the domain of ICs to incidental learning while excluding intentional group judgments elicited by active hypothesis testing and targets high in entitativity (McConnell, Sherman, & Hamilton, 1997; Srull & Wyer, 1989).

Finally, we find evidence for ICs when the stereotypes are learned in an interactive process of expressing stereotypic expectancies and receiving feedback that is met with benefits and costs (Denrell, 2005; Fazio et al., 2004). In a monetarily reinforced operant-learning variant, participants continuously predicted whether they expected positive or negative behavior of majority and minority group members. Using this active procedure, we did not only replicate ICs in end-of-sequence attitudes but also found evidence for a novel manifestation. In line with recent evidence from a non-social paradigm (Kutzner, Freytag, Vogel, & Fiedler, 2008) and closely corresponding to a prominent definition of stereotypes (McCauley & Stitt, 1978, p. 929), participants' probabilistic predictions continuously differentiated between the groups.

Together, our theoretical arguments, simulations and empirical evidence expand the applicability of ICs in reality. Contrary to the implications of previous research, even long-standing minorities for which abundant experiences exist and that have been target of active attempts to form impressions or even of discrimination seem not to be immune to illusory stereotypes.

Acknowledgments

I would like to thank Livia Keller and Celina Kacperski for very helpful comments on earlier versions of this manuscript.

References

Allan, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin, 114*, 435–448.

Chun, W., & Lee, H. (1999). Effects of the difference in the amount of group preferential information on illusory correlation. *Personality and Social Psychology Bulletin, 25*, 1463–1475.

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

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review, 112*, 951–978.

Eder, A., Fiedler, K., & Hamm-Eder, S. (2011). Illusory correlations revisited: The role of pseudocontingencies and working memory. *Quarterly Journal of Experimental Psychology, 64*, 517–532.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87*, 293–311.

Fiedler, K. (2000). Illusory correlations: A simple associative algorithm provides a convergent account of seemingly divergent paradigms. *Review of General Psychology, 4*, 25–58.

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

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom – A cognitive-environmental approach. *Organizational Behavior and Human Decision Processes, 88*, 527–561.

Fiedler, K., Walther, E., & Nickel, S. (1999). The auto-verification of social hypotheses: Stereotyping and the power of sample size. *Journal of Personality and Social Psychology, 77*, 5–18.

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., Katz, L. B., & Leirer, V. O. (1980). Cognitive representation of personality impressions: Organizational processes in first impression formation. *Journal of Personality and Social Psychology, 39*, 1050–1063.

Hamilton, D. L., & Sherman, S. J. (1989). Illusory correlations: Implications for stereotype theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 59–82). New York: Springer.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93*, 258–268.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology, 68*, 565–579.

Kutzner, F., Freytag, P., Vogel, T., & Fiedler, K. (2008). Base-rate neglect as a function of base rates in probabilistic contin-

gency learning. *Journal of the Experimental Analysis of Behavior, 90*, 23–32.
Lichtenstein, M., & Srull, T. K. (1987). Processing objectives as a determinant of the relationship between recall and judgment. *Journal of Experimental Social Psychology, 23*, 93–118.
McCauley, C., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology, 36*, 929–940.
McClelland, J. L., & Rumelhart, D. E. (1988). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. Cambridge, MA, USA: The MIT Press Computational models of cognition and perception.
McConnell, A. R. (2001). Implicit theories: Consequences for social judgments of individuals. *Journal of Experimental Social Psychology, 37*, 215–227.
McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). 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. (1997). Target entitativity: Implications for information processing about individual and group targets. *Journal of Personality and Social Psychology, 72*, 750–762.
Meiser, T. (2003). Effects of processing strategy on episodic memory and contingency learning in group stereotype formation. *Social Cognition, 21*, 121–156.
Meiser, T., & Hewstone, M. (2001). Crossed categorization effects on the formation of illusory correlations. *European Journal of Social Psychology, 31*, 443–466.
Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183–198.
Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology, 29*, 11–27.
Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1*, 33–60.
Pryor, J. B. (1986). The influence of different encoding sets upon the formation of illusory correlations and group impressions. *Personality and Social Psychology Bulletin, 12*, 216–226.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
Sanbonmatsu, D. M., Sherman, S. J., & Hamilton, D. L. (1987). Illusory correlation in the perception of individuals and groups. *Social Cognition, 5*, 1–25.
Shavitt, S., Sanbonmatsu, D. M., Smittipatana, S., & Posavac, S. S. (1999). Broadening the conditions for illusory correlation formation: Implications for judging minority groups. *Basic and Applied Social Psychology, 21*, 263–279.
Sherif, M. (1967). *Group conflict and co-operation*. London: Routledge & Kegan Paul.
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*, 305–323.
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). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36*, 1202–1212.
Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 440–463.
Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96*, 58–83.
Van Overwalle, F., & Siebler, F. (2005). A connectionist model of attitude formation and change. *Personality and Social Psychology Review, 9*, 231–274.
Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 110*, 536–563.

Received September 23, 2010
Revision received January 26, 2011
Accepted January 31, 2011
Published online May 18, 2011

Florian Kutzner

Department of Psychology
University of Heidelberg
Hauptstrasse 47-51
69117 Heidelberg
Germany
Tel. +49 6221-547366
Fax +49 6221-547745
E-mail florian.kutzner@psychologie.uni-heidelberg.de

Appendix

Example of modeling ICs within a distributed memory framework (Smith, 1990).

The 16 vertical memory traces are generated from all four possible combinations of the ideal probes at left representing majority (Maj.), minority (Min.), positive (Pos.) and negative (Neg.) valence. Subsequently, noise affects memory by rendering every feature zero with a chance of 30% (26% in this example). To obtain memory echoes for the groups, in a first step every trace's similarity with a group, for example, the minority, is determined by the sum of the feature-products (see bottom line). In a second step, every trace is weighted by its similarity with the respective group (see Example Trace). Finally, all weighted traces are summed to obtain the memory echo for the group. A group evaluation can be determined by correlating the valence segment of the group echo with the ideal probe for positive valence. Note that this example reflects an IC in that the correlation between the majority echo and positive (r = .96) is higher than between the minority echo and positive (r = .79).

Trace segments	Ideal probes				Matrix of memory traces																Example trace			Echo for ...	
	Maj.	Min.	Pos.	Neg.	Maj. pos. (9)									Maj. neg. (3)			Min. pos. (3)			Min. neg. (1)	(#7)			Maj.	Min.
					#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	Original	Weighted by similarity with majority	**Weighted by similarity with minority**		
Group	−1	1	0	0	−1	−1	−1	0	−1	−1	**−1**	−1	0	0	0	0	1	0	0	1	−1	−4	**2**	−0.08	0.01
	1	−1	0	0	0	1	1	1	1	0	**1**	1	0	1	1	0	0	−1	0	−1	1	4	**−2**	0.07	−0.01
	1	−1	0	0	1	0	1	1	1	1	**1**	0	1	1	1	0	0	−1	−1	−1	1	4	**−2**	0.1	−0.01
	−1	−1	0	0	−1	0	−1	−1	−1	−1	**0**	−1	−1	0	0	−1	−1	−1	0	0	0	0	**0**	−0.09	−0.02
	−1	−1	0	0	−1	−1	0	0	−1	−1	**−1**	−1	−1	−1	−1	−1	−1	0	−1	0	−1	−4	**2**	−0.08	−0.02
	1	1	0	0	1	0	1	1	1	1	**0**	1	0	0	0	1	1	1	1	0	0	0	**0**	0.09	0.03
Specific group member	0	0	0	0	0	−1	0	−1	−1	1	**1**	−1	1	1	0	−1	−1	0	−1	−1	1	4	**−2**	−0.01	−0.02
	0	0	0	0	−1	1	−1	0	−1	0	**−1**	1	1	0	1	0	−1	0	−1	1	−1	−4	**2**	−0.05	−0.01
	0	0	0	0	−1	−1	0	1	0	1	**1**	1	1	0	1	−1	1	1	1	1	1	4	**−2**	0.02	0.02
	0	0	0	0	−1	−1	0	0	−1	−1	**1**	1	−1	−1	0	−1	1	0	1	−1	1	4	**−2**	−0.04	0.01
	0	0	0	0	1	−1	−1	−1	1	−1	**1**	1	0	0	−1	0	1	0	0	−1	1	4	**−2**	0.01	0.01
	0	0	0	0	0	1	1	0	−1	−1	**1**	−1	−1	0	−1	−1	0	−1	1	1	1	4	**−2**	−0.02	−0.01
Valence	0	0	−1	1	0	−1	−1	−1	−1	−1	**−1**	−1	0	1	1	1	0	−1	0	1	−1	−4	**2**	−0.07	0
	0	0	−1	1	−1	0	−1	−1	−1	−1	**−1**	−1	0	0	1	1	0	0	−1	1	−1	−4	**2**	−0.09	0
	0	0	1	−1	0	1	1	1	0	1	**1**	1	1	0	0	−1	1	1	1	−1	1	4	**−2**	0.06	0.02
	0	0	−1	1	−1	0	−1	0	−1	−1	**0**	−1	−1	1	0	1	−1	−1	−1	1	0	0	**0**	−0.07	−0.02
	0	0	1	−1	0	1	1	0	1	1	**1**	1	1	−1	0	−1	1	0	1	−1	1	4	**−2**	0.06	0.01
	0	0	1	−1	1	1	1	1	1	1	**1**	1	1	−1	−1	0	0	1	0	0	1	4	**−2**	0.09	0.01
	0	0	−1	1	0	−1	0	−1	0	0	**−1**	−1	−1	1	1	1	0	−1	−1	1	−1	−4	**2**	−0.02	−0.01
Specific behavior	0	0	0	0	−1	1	−1	0	−1	−1	**0**	1	1	−1	−1	−1	1	1	0	0	0	0	**0**	−0.06	0.02
	0	0	0	0	1	−1	−1	0	1	−1	**−1**	−1	−1	1	−1	0	−1	1	1	−1	−1	−4	**2**	−0.01	0
	0	0	0	0	−1	1	0	−1	1	−1	**0**	1	1	0	−1	0	0	−1	0	−1	0	0	**0**	−0.01	−0.01
	0	0	0	0	0	0	1	−1	−1	0	**−1**	−1	−1	1	−1	−1	0	1	−1	1	−1	−4	**2**	−0.03	0
	0	0	0	0	1	−1	0	1	1	−1	**0**	−1	0	1	1	1	1	1	0	0	0	0	**0**	0.04	0.02
	0	0	0	0	−1	0	0	0	−1	−1	**0**	−1	−1	1	1	−1	0	−1	1	−1	0	0	**0**	−0.05	−0.01
	0	0	0	0	−1	−1	−1	1	−1	0	**0**	1	−1	−1	−1	1	1	1	1	−1	0	0	**0**	−0.04	0.02
	Similarity with minority				1	−1	−1	0	0	1	**−2**	1	1	−1	−1	3	4	4	3	3	Evaluation (i.e., similarity with pos.)	6	24	**−12**	Corr. with pos.	0.96	0.79
	Similarity with majority				5	3	5	4	6	5	**4**	5	3	3	3	3	2	0	1	−3					

Note. In the model and the simulations similarity values are standardized with the number of non-zero features of the trace before the traces are weighted.