Implicit learning in attractiveness evaluation:
The role of conformity and analytical processing

Ivan I. Ivanchei[a,c,f], Nadezhda V. Moroshkina[a,b], Roman V. Tikhonov[a,d], Irina V. Ovchinnikova[a,e]

[a] - Department of Psychology, Saint Petersburg State University, Saint Petersburg, Russia; [b] - Laboratory for Cognitive Studies, Saint Petersburg State University, Saint Petersburg, Russia; [c] - Cognitive Research Lab, Russian Academy of National Economy and Public Administration, Moscow, Russia; [d] - National Research University Higher School of Economics, Saint Petersburg, Russia; [e]- University of Houston, Houston, TX, USA; [f]- Ghent University, Ghent, Belgium

Author Note

This study has been supported by the Russian Foundation for Humanities (Project No. 15-36-01355).

The data reported in this paper is available at https://osf.io/82jh9.

The data was partially presented on the local scientific conference at the Department of Psychology, Saint Petersburg State University, in October 2017.

Correspondence should be addressed to: Ivan I. Ivanchei, Room 2402, Department of Psychology, Russian Academy of National Economy and Public Administration, pr. Vernadskogo, 84, korp. 9, 119571, Moscow, Russia. Email: ivanchey-ii@ranepa.ru. Tel.: +7 (921) 630-56-15.

Word count: ~ 9035

Abstract

We address the question how people's opinion and features of information interact in the process of indirect social influence. Implicit learning was considered as a mechanism for conformity in social perception. We carried out two experiments using a hidden covariation detection paradigm. In a learning phase, participants memorized a set of female photographs presented together with their attractiveness ratings. The ratings correlated with the hairstyle of the photographed women. The participants who did not consciously detect this correlation demonstrated a systematic bias towards the correlation when evaluating the new stimulus persons. Information about the source of the ratings in the learning phase (other people's opinions or non-social sources) did not modulate learning. Learning was not observed when participants critically evaluated the ratings during the memorisation phase. The study shows that (i) conformity may be based not only on reinforcement learning mechanism (as was previously suggested) but also on unsupervised implicit learning; (ii) implicit learning occurs automatically irrespective of the context (social or not); and (iii) a critical attitude towards learned material may prevent implicit learning from being manifested in a test phase. We conclude that indirect social influence may be affected by people's opinion towards the provided information. The study makes contribution to both implicit learning and social perception research.

Keywords: implicit learning, face perception, facial attractiveness, conformity, decision strategy

Implicit learning in attractiveness evaluation:
The role of conformity and analytical processing

People are influenced by numerous information streams that flow through TV, newspapers, social media, personal communication, and other channels. Such influences can be considered from different perspectives. On one hand, we can look at the properties of the information provided (e.g., the presence of evidence, internal congruency) and characteristics of its source (e.g., its credibility). On the other hand, we can consider individual features of the recipient, (e.g., her initial attitudes, or experience). In this study, we investigate the interaction between these factors in a situation of indirect social influence.

From previous studies, we know that in the absence of their own opinion or relevant knowledge, a person's decisions can be influenced by knowingly irrelevant information (anchoring effect, Tversky & Kahneman, 1974). If one has an opinion and it contradicts the presented information, decision-making will depend on the source of the information provided. Research shows that the information from a social source, i.e. a group opinion, influences our decisions most strongly, because apart from the information influence, it has a normative value. In the case of informational social influence, a person follows the content of the information provided by a group in an attempt to find the correct decision. In the case of normative social influence, the feeling of being a part of a group is a dominant motivation (Deutsch & Gerard, 1955; Cialdini & Goldstein, 2004). However, traditional conformity research (e.g., Asch, 1951) usually addresses a change of opinion regarding particular stimuli which were directly judged by a group. In contrast, we are interested in an indirect social influence.

In our study, we modelled the indirect social influence with the hidden covariation detection (HCD) which is an example of an unconscious information influence. HCD is an experimental effect which is usually observed in the situations where a person gets familiarized with a set of stimuli containing a hidden covariation then evaluates new stimuli. A series of studies showed that a person unconsciously shifts their evaluation in accordance with the imposed covariation (Lewicki, 1986a, 1986b; Hill, Lewicki, Czyzewska, & Boss, 1989). HCD is an example of implicit learning, which started being investigated in the late sixties (Reber, 1967). Implicit learning is the phenomenon of unconscious processing of regularities in the environment. It was shown to be relevant in various domains including language learning (Perruchet, 2008; Winter & Reber, 1994), motor learning (Nissen & Bullemer, 1987), category formation (Waldron & Ashby, 2001). However, the HCD phenomenon was one of the most controversial. Some authors claimed its ubiquity (Lewicki et al., 1997), while others treated it as an artifact (Hendrickx, De Houwer, Baeyens, Eelen, & Van Avermaet, 1997).

Presented below is a brief overview of the main findings in the area of implicit learning, especially HCD with an ecologically valid task (e.g., first impression formation based on facial appearance). We focus on successful and unsuccessful replications of the HCD effect and discuss the participants' attitudes as a factor that may resolve these contradictions. Next, we consider social influence studies which demonstrate that information from a social source affects individual judgments to a greater extent than socially irrelevant information. We present two experiments investigating the role of participants' agreement with the presented information on the implicit learning of hidden regularities contained in this information.

Implicit learning

HCD has been one of the most intriguing research areas since the first studies by Lewicki and colleagues (Lewicki, 1986a, 1986b; Hill, Lewicki, Czyzewska, & Boss, 1989; Hill, Lewicki, Czyzewska, & Schuller, 1990). These studies showed that participants could unconsciously learn hidden covariations between facial features (e.g., type of hairstyle) and psychological features (e.g., intelligence, righteousness, or kindness). Usually, Lewicki's HCD paradigm consists of three phases: In the first phase (learning phase), participants are presented with stimuli belonging to one of two different categories. For example, in Lewicki's seminal experiment (Lewicki, 1986a), women in the photos were described either as kind or capable. Unbeknownst to participants, this feature was covaried with the haircut of the women in the photo (long or short). In the second phase (test phase), the participants had to categorize new pictures as belonging to one of the two categories. In the third phase (awareness evaluation), participants completed a questionnaire to ascertain their awareness of the critical covariation. The result in most of Lewicki's studies was that participants' responses were influenced by the induced covariation, even if they were not aware of this covariation.

These results were striking because they indicated that important aspects of social perception can be easily influenced by an irrelevant piece of information accidentally correlated with psychological traits. Moreover, Lewicki and colleagues showed that once aware of a hidden covariation, participants had a tendency to use it more (Hill, Lewicki, Czyzewska, & Schuller, 1990). These studies influenced the development of Social Perception research. For example, Zebrowitz considered the HCD effect as a manifestation of the same mechanism which is responsible for the effect of familiar face overgeneralization, i.e. transfer of known psychological features to persons with familiar faces (Zebrowitz, 1996; Zebrowitz & Collins, 1997).

Hendrickx et al. (1997) tried to replicate 12 experiments from Lewicki and colleagues' papers but succeeded only once. Hendrickx et al.'s (1997) criticism of the HCD paradigm was primarily methodological: primarily poor methods description, control of awareness, and control of potential confounds (features partly correlated with crucial "hidden" features). However, the response of Lewicki et al. (1997) was more theoretical. They argued that Hendrickx et al. (1997) were inaccurate in their replication attempts. In particular, a larger number of aware participants in Hendrickx et al. (1997) study was explained not by more sensitive awareness measures, but by the induction of analytical processing strategies in the participants.

As a result of this fundamental disagreement, there were several attempts to replicate the HCD effect by other researchers in this field. Some were successful (Barker, Andrade, 2006; Karpov, Moroshkina, 2014; Moroshkina, Karpov, 2015), other were not (Bröder et al., 2007). To resolve this contradiction, some studies focused on the strategy which participants use in implicit learning. Karpov and Moroshkina (2014) modified Lewicki's (1986a) paradigm by using one bipolar scale of intelligence (low – high intelligence) in the test phase. After the test, they asked participants to explain their decisions. They not only replicated the HCD effect, but also observed a negative correlation between a quantity of participants' verbal explanation of their decisions and the consistency of these decisions with the hidden covariation. That is, the decisions with more extensive verbal explanations, were rarely in accordance with the hidden covariation. These results point to the possibility of two different types of decision-making in HCD experiments based on either existent explicit or acquired implicit knowledge. Stamov Rossnagel (2001) used the brain scans paradigm (Lewicki, Hill,

& Sasaki, 1989) to successfully replicate the original results. However, they also showed that HCD effect does not occur with explicit instruction (when the participants are encouraged to look for the critical feature in stimuli) and analytical processing (forcing participants to focus on parts of the stimuli and not on the stimuli as a whole).

In other areas of the implicit learning research, a similar pattern of results was demonstrated: the implicit learning effect disappeared or was reduced when participants were using an analytical (active) strategy, but the effect was robust when the participants' strategy was intuitive (passive). This effect was found in the contextual cueing paradigm (Lleras & Von Mühlenen, 2004) and in the artificial grammar learning (Kinder et al., 2003). In accordance with these results, Ivanchei and Moroshkina (2018) observed a more analytical process in the artificial grammar learning when additional metacognitive reports were used in the test phase. Concurrent verbalization can also influence implicit knowledge manifestation, which was observed in various paradigms (Moroshkina, Ivanchei, Karpov, & Ovchinnikova, 2019). The main conclusion from these studies was that participants may adopt different decision-making strategies, some of which may prevent hidden covariation processing or knowledge expression. This conclusion is common for different subfields of implicit learning including HCD, contextual cueing, and artificial grammar learning.

The above examples support the idea that participants may adopt different decision-making strategies (holistic or analytical) which in turn may affect the manifestation of implicit learning. By the holistic strategy we mean the reliance on the first impression of a stimulus in its evaluation. Such decisions are usually difficult to explain verbally and are often attributed to intuition or a gut-feeling. Analytical strategy, in turn, is based on exact features of a stimulus and the decisions can be explained verbally. We suppose that participants' attitude towards the presented information may be critical for the adoption of a particular strategy in the experiment. By the attitude in this context we mean a general agreement or disagreement of a participant with a direct (explicit) part of the information provided. In particular, a disagreement with the presented information is expected to lead to the analytical strategy. However, a contradiction between the learned regularities and prior knowledge or participants' attitudes to the stimulus was never directly addressed in the HCD and implicit learning studies. This was one of the points in Lewicki et al.'s (1997) response to Hendricx et al. (1997). They argued that while preparing their experiments, they tried ensuring that participants see the task as meaningful and do not try “to outsmart the experimenter” (Lewicki et al., 1997, P. 223). Otherwise, participants may start to behave analytically, which “interferes with the natural mode of encoding of the stimuli” (Lewicki et al., 1997, P. 223).

Social influence

Conformity studies have a long history. While classic studies demonstrated a social influence on the evaluation of socially irrelevant objects (e.g., Asch, 1951), several recent studies showed a shift of facial attractiveness evaluation under a group opinion (see e.g., Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Zaki, Schirmer, & Mitchell, 2011). In Klucharev et al. (2009), participants rated the attractiveness of presented faces then observed average ratings of a large group of people, which had been manipulated by the experimenters. Critically, on the repeated assessment of the same set of photos, participants shifted their evaluations towards the “group opinion”. Moreover, Klucharev et al. (2009) found the brain activity associated with conformity trials which indicated that conflict with the group opinion was processed as error signal in reinforcement learning. Both the behavioral and fMRI results were much weaker in the control experiment, where participants compared their ratings with “computer-generated” ratings. That is, conflict with the group opinion lead to a stronger correction due to the normative value of the group opinion.

Some authors addressed the possibility that participants’ responses may shift towards the group opinion not because of a normative influence of the group opinion but just as a result of perceptual priming or anchoring. Germar et al. (2014) showed that in a simple color discrimination task, other people’s responses presented to participants affected their decisions even when they were informed that these decisions came from a completely different task. Similar results with another perceptual decision-making task were obtained by Eskenazi et al. (2016). Using the gaze-cueing paradigm, they found that irrelevant gaze directions presented to participants influenced their confidence judgements on the primary task decisions even though the participants were instructed to ignore these stimuli. However, this effect did not work with non-social stimuli like cars oriented right or left. These studies show that information that is explicitly irrelevant to the main task may affect participants’ decisions. However, this effect was much stronger when the participants believed they observed relevant information about the opinions of others. This suggests that informational or normative social influence might be additive to the simple anchoring effect.

However, some studies suggest that social stimuli may be specific. Gamond et al. (2012) used two learning procedures associating some facial features with socially relevant (flexible or determined personality) and irrelevant (abstract category “A” or “B”) traits. In the first procedure, participants had to categorize faces by themselves, then received feedback on the accuracy of their responses. In the second procedure, participants just passively observed the covariation between facial features and social or non-social traits. Learning was found for both social and non-social features using the second procedure. The first procedure was effective only for association learning of non-social features. Gamond et al. (2012) made a conclusion that social category learning has a different nature than non-social category learning. They also noted that participants might have some prior knowledge about the psychological traits which interfered with the associations induced in the laboratory experiments. This may be particularly salient with a trial-by-trial feedback procedure.

Gamond’s et al. (2012) results contradict the results listed above. We assume, however, that in all the discussed studies, the critical variable was participants’ attitude towards the presented material. Based on the review of implicit learning, we suppose that this variable mediates the influence of social context. That is, in a social condition, participants tend to agree with the presented material more than in the non-social conditions.

In the present study, we varied the narrative about the source of provided information to investigate the influence of participants’ attitude towards the information on HCD. We had three conditions with different narratives, which were aimed to vary the extent to which provided information may cause participants’ disagreement. Instead of psychological features, we used facial attractiveness covaried with facial features. This variable is well suited for our aims. First, the evaluation of others’ appearance is automatic, fast, and consistent across subjects (Willis & Todorov, 2006). Second, most people have their own criteria for facial attractiveness, thus we have a high chance to provoke a disagreement with the stimulus material. Third, society has strong standards of facial attractiveness influencing people’s decisions, which increased the ecological validity of the study.

Overview of the experiments

For the present study, we used a specially created photo database of female faces with varying hairstyles (described further in the Materials section). The HCD paradigm was based on that used by Moroshkina and Karpov (2014) and Karpov and Moroshkina (2015). This implementation of the HCD paradigm has several improvements compared to that of Lewicki. First, it uses a counterbalancing scheme, i.e. the same face with different hairstyles was presented to different participants to introduce the opposite covariations (see Design and procedure section in Experiment 1). Second, it is followed by a detailed, funnel-like awareness assessment questionnaire, as suggested by Hendrickx et al. (1997).

Two experiments were conducted. Participants memorized the faces and the ratings of attractiveness associated with them. We varied the narrative about the source of these ratings (social / non-social / random) in order to explore the role of the participants' attitude towards the information in the stimulus material. Based on the normative social influence approach, we predicted that in the social narrative condition, participants will agree with the imposed information more than in the non-social condition. Next, based on the idea of participants' attitude as a factor of decision-making, we predicted that the HCD effect will be larger in the social condition compared to the non-social condition. The random narrative condition was predicted to have the smallest HCD effect based on the pure anchoring effect (i.e. the association between the numbers and the perceptual features of the stimuli). The second experiment was designed to clarify the findings of the first experiment.

Experiment 1

For the first experiment, we developed a paradigm that was similar to the one used by Lewicki (1986; see also Moroshkina & Karpov, 2015) to investigate the implicit learning in the attractiveness evaluation. We have introduced a hidden covariation in the learning material (the association between the attractiveness ratings and the stimulus person's hairstyle) and we expected participants to bias their own attractiveness ratings for the test photos in accordance with this hidden regularity. Additionally, we investigated the influence of the narrative about the source of the learned information on implicit learning.

Method

Participants

Eighty-eight volunteers (57 women, 18–39 years old, mean age = 25.5) took part in the experiment. They were not paid for their participation. Participants were current undergraduate or master students, or already completed their degrees. All participants gave informed consent before participation. The study was approved by the Ethics Committee of St. Petersburg State University. The sample size was determined by our previous studies with this paradigm (Moroshkina & Karpov, 2015).

Materials

We used 68 female photos for the experiment. They were selected from the Russian Database of Neutral and Smiling Female Faces (Moroshkina et al., 2018). The stimuli from the database are female portrait photos with the models' own choice of clothes and makeup. The models were 17-27 years old and had Caucasian appearance. We used photos with neutral faces. For the learning phase, a set two sets of 20 photos was used. There were two sets of photos for the learning phase for balancing purposes (see Design and Procedure section). Another set of 28 photos was used for the test phase. The photos were 11×15 cm in size. Participants sat at a distance of approximately 60 cm from a 24-inch computer monitor. The experiment was run using PsychoPy 1.84.2 (Peirce, 2007).

Design and procedure

The participants were told that they were taking part in a study of a first impression. In the learning phase, they were presented with 20 female photographs one at a time. Each photograph contained a rating on a 10-point attractiveness scale (where 1 = *low attractiveness* and 10 = *high attractiveness*). The task for the participants was to try to memorize all faces that had a rating above 5.

The 88 participants were randomly assigned to one of three groups. Narratives about the source of the presented ratings differed for each group during the learning phase. In the first group, participants were told the ratings reflected "the opinions of other people on the attractiveness of the faces" (*social* narrative). Participants of the second group were told the attractiveness ratings were obtained by “the objective analysis of the relative sizes and the positions of facial features” (*computer* narrative). The third group of participants were told “the numbers on the rating scale were chosen randomly” (*random* narrative). These explanations were not true. We debriefed the participants regarding the deception at the end of the study.

Unbeknownst to the participants, the ratings were correlated with the stimulus person’s hairstyle. Half of the participants in each group were presented with high attractiveness ratings (from 6 to 10) for the loose-haired women and low attractiveness rating (from 1 to 4) for the gathered-haired women, while the other half observed the opposite covariation. In both subsets, high ratings were assigned to the same stimulus persons with different hairstyles. Thus, for the learning phase, we used two different photos of the same stimulus person. However, each participant was presented with only one photo of each stimulus person – either with loose or gathered hair.

During the learning phase, the first and second groups were asked to draw a vertical bar on a piece of paper each time they disagreed with the attractiveness ratings they saw on the screen.

During the test phase, participants were presented with 28 photos of new stimulus persons. The task was to rate the attractiveness of the women using the same 10-point scale (1 = *low attractiveness* and 10 = *high attractiveness*). Participants used a computer mouse to give their rating answers. During the test phase, one half of the stimulus models had loose hair and the other half had gathered hair. The order of the stimuli was randomized for every participant.

At the end of the experiment, participants were given a questionnaire to assess their awareness of the hidden regularity (the correlation between the ratings and the hairstyle) and their general feelings about the learning phase ratings (mostly agree or disagree). See Appendix for the exact formulation of the questions.

Thus, in the first experiment, we had three independent variables: source narrative (three levels: *social*, *computer*, or *random*), covariation type (two levels: loose or gathered hairstyle associated with high ratings in the learning phase), and stimulus person’s hairstyle in the test phase (two levels: loose or gathered hairstyle). We anticipated that participants might learn the hidden covariation between the type of hairstyle and the attractiveness ratings, and generalize it to the new faces. For example, the participants who observed higher ratings for the loose-haired women during the learning phase might rate the loose-haired women during the test phase higher than those with gathered hairs and vice versa. According to the social normative influence, we expected the highest level of agreement with the presented material in the *social* condition. Thus, we expected the higher level of learning in this condition.

To recap, we expected no or a very weak HCD effect in the *random* condition (weak learning may be possible due to the anchoring effect). In the *computer* condition we expected HCD due to the informational influence, and in the *social* condition we expected the largest HCD due to a combination of information and normative social influences reducing participants' disagreement with the stimulus material.

Results

Based on the post-experimental interview, we excluded six participants from the dataset who mentioned the hairstyle in their responses to *Q1* or *Q3* (see Appendix). Two of them were from the *social* group and four from the *random* group.

The level of disagreement with the models' ratings

In the post-experiment interview data, we found that in the *social* and *computer* conditions, more participants responded that they disagreed in general with the ratings in the learning phase, while in the *random* condition, more participants agreed with these ratings or responded that they did not think of it (see Table 1). We combined the responses "Yes, the ratings were congruent with my own impressions" and "I had not thought of it until you asked me about it" into the same category (yes) for the analysis.

Table 1. The number of participants in Experiment 1 with different responses to *Q2* in the post-experimental interview ("Did the attractiveness ratings seem plausible while you were memorizing the women's faces during the first stage?")

	Narrative			
Response	*Social*	*Computer*	*Random*	Total
"Yes"	8	12	16	36

"No"	18	19	9	46
Total	26	31	25	82

The Chi-squared test indicated a significant difference between the distributions of *yes* and *no* responses in the three groups, $\chi^2(2) = 6.26$, $p = .044$. As we can see from Table 1, in the *random* condition, more people responded "Yes" to Q2, and in the *social* and *computer* conditions, more people responded "No".

We also evaluated the concurrent disagreement with the attractiveness ratings in the *social* and *computer* conditions and their relation to the post-experimental disagreement. The *social* narrative participants marked on average 5.8 (out of 20; *SD* = 2.4) photos whose ratings they did not accept. In the *computer* condition, the average number of disagreements was 5.5 (*SD* = 2.8). This difference did not reach significant levels, $t(52) = 0.41$, $p = .683$. The concurrent disagreement during the learning phase was related to the disagreement reported in the post-experimental interview, $t(28.4) = -2.21$, $p = .036$. Those participants who reported that the attractiveness' ratings seemed implausible to them, reported disagreement with the photos rating an average of 6.2 (*SD* = 2.1) times in the learning phase. Those who responded that the attractiveness ratings were plausible or who did not think about their plausibility reported disagreement with the photo's rating an average of 4.5 (*SD* = 3.0) times.

HCD effect

To account for both between-subject and between-stimuli variance, we applied a linear mixed effects model with random intercepts for each participant and each stimulus using 'lme4' package for R (Bates, Maechler, Bolker, & Walker, 2014). The P-values were computed using Kenward-Roger approximation for the degrees of freedom implemented in 'lmerTest' package (Kuznetsova, Brockhoff, & Christensen, 2017). We also ran Bayesian analysis to evaluate the null results.

The results are presented in Figure 1. We observed significant learning only in the *random* condition, as revealed by the Hairstyle × Learning subset interaction, $F(1, 647) = 13.89$, $p < .001$. This interaction indicates that participants evaluated the stimulus persons during the test phase in accordance with the learning phase covariation. Those who observed the association between the high ratings and loose hair in the learning phase evaluated the loose-haired women in the test phase higher than the gathered-haired women and vice versa. In the *social* and the *computer* conditions, the interaction was not significant, $F(1, 647) = 0.40$, $p = .527$, and $F(1, 809) = 1.25$, $p = .264$, respectively. The full model with the narrative as an additional factor showed a marginally significant three-way (Hairstyle × Learning subset × Narrative) interaction, $p = .077$. *Random – social* narrative contrast was significant ($p = .039$), indicating differences between the conditions in implicit learning. The *random – computer* narrative contrast was marginally significant ($p = .060$). The *social – computer* narrative contrast was not significant ($p = .778$).

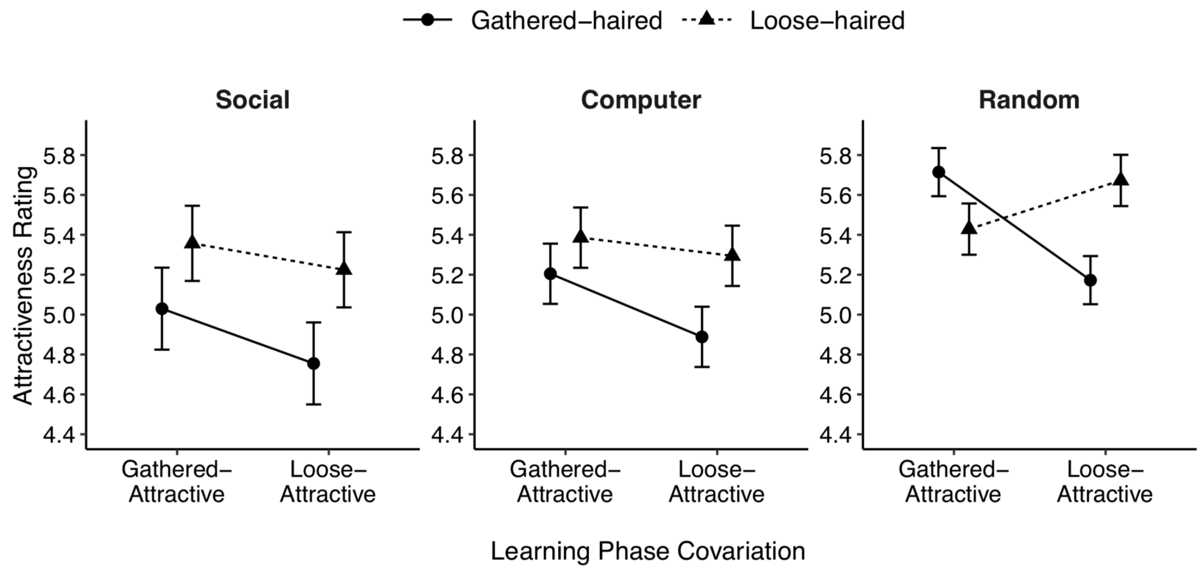

Figure 1. Mean ratings for loose-haired and gathered-haired stimulus persons by two covariation type subgroups in three narrative conditions in Experiment 1. The error bars correspond to 95% confidence intervals.

We also performed a Bayesian analysis to test the evidence for or against the Hairstyle × Learning subset interaction in the three groups. To set informed priors, we ran the mixed effects model on the data from Moroshkina & Karpov (2014) and computed the standardized effect size for the interaction – Fisher z transformed *Pearson's r* = 0.14. Note that effect size is small due to the high number of *df* approximated by the Kenward-Roger method (Kuznetsova, Brockhoff, & Christensen, 2017). We used a half-normal distribution with M = 0 and SD = 0.14 as a prior for the alternative hypothesis. In the *social* condition, *BF10* for the interaction effect was 0.46. In the *computer* condition, *BF10* for the interaction effect was 0.74. Both Bayes factors indicate evidence against the interaction effect, however, they do not reach the conventional cut-off level for substantial evidence for the null (*BF10* = 0.33). In the random condition, the *BF10* was 357.87 indicating substantial evidence for the interaction model, thus supporting the traditional analysis presented above.

The relationships between disagreement and HCD effect

After evaluating HCD in each of the three conditions, we decided to combine the data on the post-experimental disagreement in the *social* and *computer* conditions as there was no learning in either of them and to compare them to data obtained in the *random* condition (see Table 1). The difference was significant, $\chi^2(1) = 4.78$; $p = .029$. Thus, those participants who did not show any learning did not agree with the ratings in the learning phase more often than those who learned the hidden covariation.

After finding the connection between learning and the amount of disagreement on the group level, we decided to look at this relationship on the individual level. Participants were divided into two groups on the basis of their answers to *Q2* in the post-experimental interview ("Did the attractiveness ratings seem plausible to you while you were memorizing the women's faces during the first stage?"). HCD in all participants based on their *Q2* reply can be seen in Figure 2. We ran the analysis of learning on the whole dataset with the additional factor of *agreement* (with the ratings in the learning phase) with two levels: *mostly agree* or *mostly disagree*. The full model for all the groups together showed a significant three-way Hairstyle × Agreement × Learning subset interaction, $F(1, 2184) = 13.09$, $p < .001$, indicating the influence of agreement on the hidden covariation learning. The bias towards the imposed regularity, indicated by the Hairstyle × Learning subset interaction was also significant for the full model, $F(1, 2184) = 11.82$, $p < .001$. A separate model for the agree-participants showed significant learning of the hidden regularity, $F(1, 944) = 24.10$, $p < .001$; a separate model for disagree-participants did not, $F(1, 1214) = 0.02$, $p = .895$. Bayesian analysis supported the interaction model for agree participants, *BF10* = 27415.88, and did not provide any conclusive evidence for disagree participants, *BF10* = 0.91.

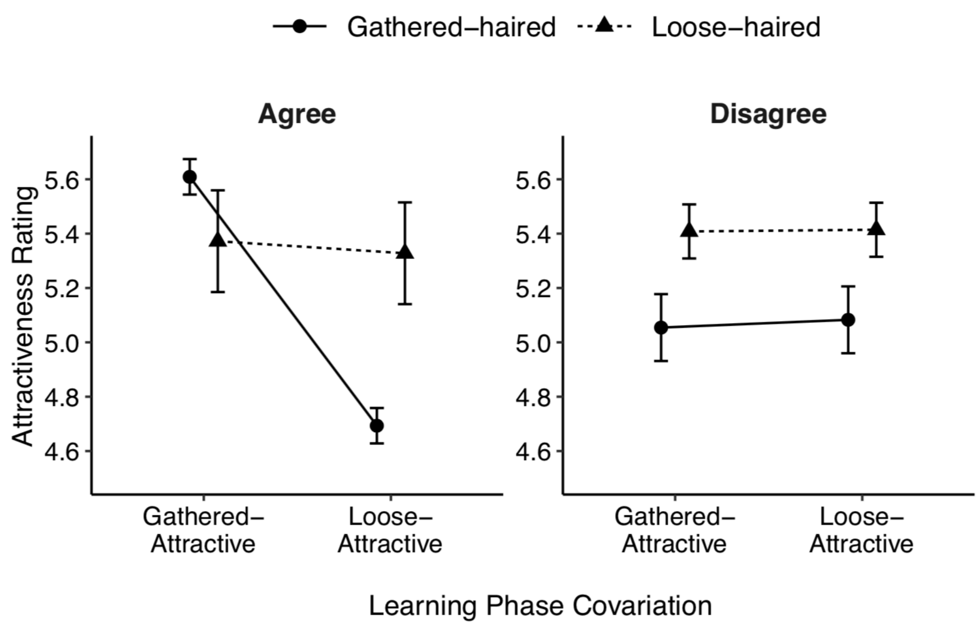

Figure 2. Mean ratings for loose-haired and gathered-haired stimulus persons by Agree and Disagree participants depending on their learning phase. The left panel represents the participants who mostly agreed with the ratings in the learning phase. The right panel represents the participants who mostly disagreed with these ratings. Error bars correspond to 95% confidence intervals.

To assess the relationships between the concurrent disagreement in the learning phase and the HCD effect in the test phase, we added a covariate to the standard model. Combined data from *social* and *computer* conditions were analyzed (*random* narrative participants did not report concurrent disagreement). A three-way Hairstyle × Learning subset × Disagreement mixed-effects model revealed a significant two-way Hairstyle × Learning subset interaction, $F(1, 1428) = 6.64$, $p = .010$, as well as a three-way Hairstyle × Learning subset × Disagreement interaction, $F(1, 1428) = 4.84$, $p = .028$. Regression coefficients inspection revealed that the two-way interaction indicates a more positive evaluation of the gathered-hair women after gathered – attractive covariation in comparison to loose – attractive covariation, which is the HCD effect. The three-way interaction coefficient is negative ($B = -0.16$, $SE = 0.07$) indicating a decrease of the HCD effect with the increase in the number of concurrent disagreement reports during the learning phase.

Discussion

We hypothesized that participants' opinions based on their previous experience, will contradict to the presented information during the learning phase, which may decrease learning. Social influence was expected to reduce the effect of disagreement, i.e. the contradictions will be resolved in line with the group opinion leading to a lesser amount of after-experimental and concurrent disagreements and as a result, increased HCD effect in the *social* condition. However, the level of both concurrent and post-experimental disagreement was approximately the same in the *social* and *computer* conditions. At the same time, the level of post-experimental disagreement in fact influenced HCD. It just appeared to be unrelated to the experimental variation.

Among conditions, HCD was significant only in the *random* condition, where the participants were told that the ratings of the women whose faces they tried to memorize were just random numbers. In this condition, the participants who observed higher ratings in loose-haired stimulus persons in the learning phase gave higher ratings to the new stimulus persons with the same type of hairstyle in the test phase. The learning seems to be implicit, because we excluded from the analysis those participants who noticed the relationship between the stimulus persons' hairstyles and their ratings in the learning phase.

It seems that the critical factor in our experiment was the fact that in *social* and *computer* conditions, participants could report disagreements with the ratings presented in the learning phase. Thinking about agreement or disagreement with the presented information forced participants to consciously think about their criteria for attractiveness evaluation and, thus, made them adopt the analytical strategy. The analytical strategy, in turn, prevented participants from either learning the hidden covariation or exhibiting their learning in the test phase.

The analysis of the post-experimental agreement with the stimulus material showed that participants in the *social* and *computer* conditions reported disagreement with the ratings more often in contrast to the *random* condition, in which the participants mostly did not think about the presented ratings. When the data was split on the basis of agreement, there was a reliable HCD effect in those participants who agreed with the ratings and no sign of HCD effect in the participants who disagreed with the ratings.

The number of disagreements during the learning phase was related to the post-experimental disagreement. It was also related to the HCD effect; a higher number of disagreements in learning phase was associated with a weaker HCD effect in the test phase. Therefore, both disagreement measures converge on their relation to the HCD.

In the second experiment, we aimed to test the hypothesis that the opportunity to report a disagreement during the course of the learning phase was the factor that prevented the participants from implicit learning of a hidden covariation. We have replicated the design and the procedure of the first experiment but removed the possibility to formally disagree with the stimulus persons' ratings in the learning phase.

Experiment 2

Method

Participants

Eighty-five participants (43 women, 17–25 years old, mean age = 20.3) were recruited for the experiment among non-psychology students of St. Petersburg State University who did not participate in Experiment 1. They were not paid for participation. All participants gave informed consent, and the study was approved by the Ethics Committee of St. Petersburg State University.

Materials

The stimulus material was the same as in Experiment 1.

Design and procedure

The procedure and the design were similar to Experiment 1. The only difference was that in the learning phase, participants did not have to express their disagreement with the attractiveness ratings. Therefore, we expected participants to be less critical towards the information presented in the learning phase.

Results

In the post-experimental interview, two participants reported that they relied on hairstyle during the learning phase and were excluded from analysis. One participant was from the *social* group, and one from the *random* group.

The level of disagreement with the models' ratings

The number of participants reporting a general agreement with the stimulus persons' ratings in the learning phase is presented in Table 2. The number of participants in three conditions differ because of the exclusion of the aware participants. Unlike the results of Experiment 1 (Table 1), there was no significant relationship between condition and agreement with the stimulus persons' ratings, both in a three-group analysis, $\chi^2(2) = 3.33$; $p = .189$, and in a two-group analysis (*social* and *computer* conditions vs. *random* condition) $\chi^2(1) = 0.34$, $p = .562$. *Social* condition vs. *computer* and *random* conditions test also did not reach significance, $\chi^2(1) = 2.55$, $p = .110$.

Table 2. The number of participants in Experiment 2 with different responses to the post-experimental question: "When you tried to memorize the models' faces in the learning phase did their ratings seem plausible to you?"

	Narrative			
Answer	*Social*	*Computer*	*Random*	Total
Yes	21	13	10	44
No	11	16	12	39
Total	32	29	22	83

HCD effect

The main aim of the second experiment was to clarify whether HCD effect occurs in the absence of the task to agree or disagree with the presented ratings. The analysis approach was the same as in the Experiment 1. The HCD effects in all groups are presented in Figure 3.

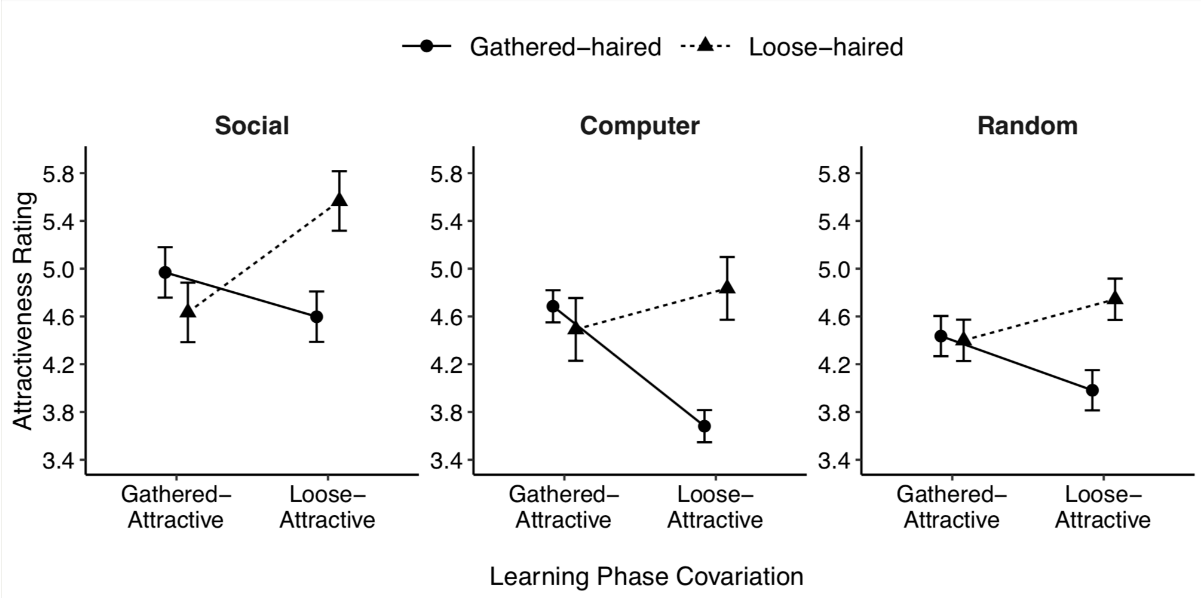

Figure 3. Mean ratings for loose-haired and gathered-haired stimulus persons by two types of covariation subgroups in three narrative conditions in Experiment 2. Error bars correspond to 95% confidence intervals.

In all the narrative conditions, a two-way mixed-effects regression with stimulus person's hairstyle and learning subset as independent variables indicated a significant bias towards the hidden regularity imposed in the learning phase by a significant interaction, $F(1, 836) = 33.08$, $p < .001$ for the *social* condition, $F(1, 783) = 32.54$, $p < .001$ for the *computer* condition, and $F(1, 566) = 9.40$, $p = .002$ for the *random* condition.

The full model with the narrative as additional factor showed no significant three-way Hairstyle × Learning subset × Narrative interaction indicating no significant differences between groups in HCD, $F(1, 2237) = 1.41$, $p = .243$. No contrasts were significant. We again performed a Bayesian analysis to test evidence for or against an interaction model in three groups. In Experiment 2, we used the size of interaction from the *random* condition in Experiment 1 (0.786 in raw units) as an SD for half-normal prior distribution. In all the groups, we obtained large Bayes factors indicating evidence for the interaction models ($BF10 > 5 \times 10^6$ in the *social* condition, $BF10 > 3 \times 10^6$ in the *computer* condition and $BF10 = 43.3$ in the *random* condition). Thus, the Bayesian analysis was consistent with the statistics of the traditional approaches reported above.

The relationships between disagreement and HCD effect

As in the first experiment, we analyzed how post-experiment agreement with the presented ratings interacted with learning. We added the agreement factor in the learning analysis. The full Hairstyle × Learning subset × Agreement model for the whole dataset provided only one significant effect: the stimulus person's hairstyle and subset interaction, indicating that there was learning of the hidden regularity, $F(1, 2239) = 71.98$, $p < .001$. The agreement factor did not interact with this effect, in contrast to the first experiment, $F(1, 2239) < 0.01$, $p = .958$, which is clearly visible in Figure 4. In accordance with the absence of significant interaction, separate models for the agree and disagree participants demonstrated significant learning effects (Hairstyle × Learning subset interactions), $F(1, 1188) = 35.54$, $p < .001$; $F(1, 1025) = 37.13$, $p < .001$, accordingly. Bayesian analysis with half-normal prior from the interaction in agree participants in Experiment 1 (SD set to 0.871) supported the interaction model for agree participants in Experiment 2, $BF10 > 10\times10^6$, and for the disagree participants in Experiment 2, $BF10 > 19\times10^6$.

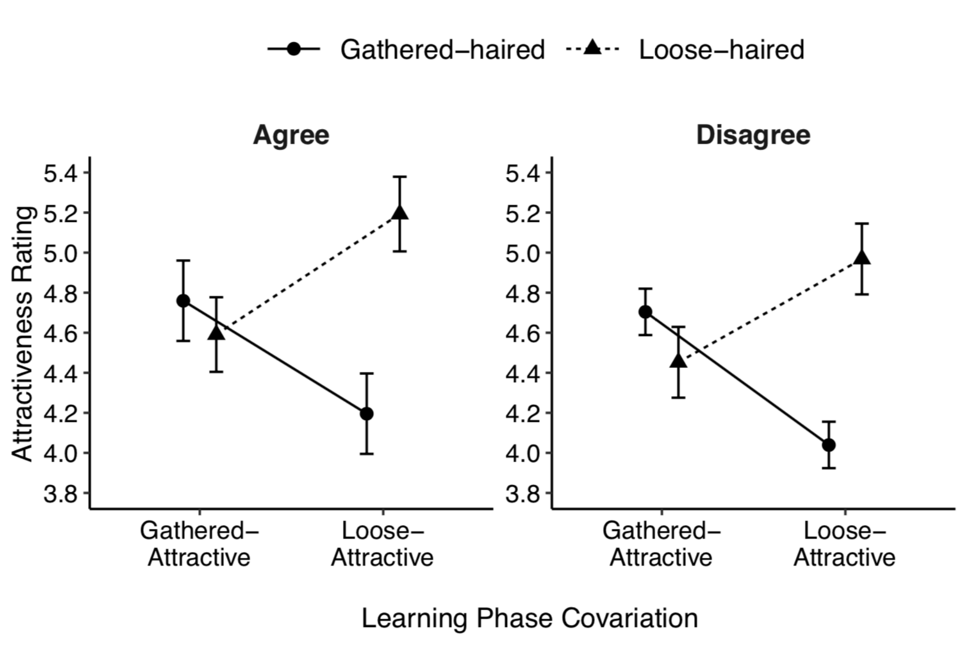

Figure 4. Mean ratings for loose-haired and gathered-haired stimulus persons by agree and disagree participants depending on their covariation type. The left panel represents the participants who (post-experimentally) mostly agreed with the attractiveness ratings in the learning phase. The right panel represents those participants who (post-experimentally) mostly disagreed with these ratings. Error bars correspond to 95% confidence intervals.

Discussion

The difference between Experiment 1 and Experiment 2 was that participants were not allowed to express their disagreement with the attractiveness ratings attached to the photographs during the learning phase. We hypothesized that this factor prevented participants from the holistic processing of stimuli and thus from exhibiting any HCD. The results of Experiment 2 supported this hypothesis. Without the possibility to actively disagree, all three groups demonstrated HCD effect. Additionally, and in contrast to the first experiment, the after-experiment disagreements with the ratings in the *social* and the *computer* conditions were not higher than in the *random* condition. Moreover, it was also not related to the learning of the hidden covariation.

General discussion

In the present study, we aimed to replicate HCD effect and investigate the factors influencing it. The results suggest that we can affect attractiveness evaluation by short-term unconscious learning. In two experiments, we observed how participants evaluated model attractiveness in accordance with the regularity implicitly imposed during the learning phase. This effect was clearly visible due to counterbalanced design of our studies. Subgroups of participants evaluated the same test photos differently depending on the arrangement of the stimuli in the learning phase. Despite the general preference for the loose-haired stimulus persons in our sample, those participants who were presented with higher ratings for gathered-haired stimulus persons than for loose-haired stimulus persons provided higher ratings for new gather-haired stimulus persons in the test phase. And vice versa: those participants who were presented with higher ratings for loose-haired women provided higher ratings to new stimulus persons with the same type of hairstyle in the test phase. Interestingly, in post-experimental interview, most participants (68% in Experiment 1 and 84% in Experiment 2) claimed to be independent of the opinions of other people.

In these experiments, we managed to demonstrate the shift in facial attractiveness evaluation after a very short learning phase (20 trials). In implicit learning studies, the effect of short-term implicit learning on liking is well established ("structural mere exposure effect"; see Zizak & Reber, 2004). However, it was typically observed on some meaningless stimuli such as a set of Latin letters. Participants usually do not have any *a priori* criteria for such stimuli preference evaluation. Hill et al. (1990) reported a shift of intelligence evaluation by photographs. As intelligence does not have any strong facial manifestations (see e.g. Kleisner, Chvátalová, & Flegr, 2014), the finding in Hill et al. can be compared to structural mere exposure results. Our study was quite different, because people usually have some *a priori* criteria for attractiveness evaluation.

There is another difference between the structural mere exposure effect and our findings. In the former case, participants demonstrate sensitivity to an A vs. not-A distinction (e.g. grammatical vs. ungrammatical stimuli) through the observation of A-stimuli. In our case, participants became sensitive to an A vs. B distinction. They were presented with the positive and negative examples (loose and gathered hairstyle) of the same regularity and with equal probability. Thus, the observed preferences shift effect cannot be attributed to mere exposure or familiarization with some prototype.

The main problem of our study was the role of participants' attitude towards the presented information. In our experiments, this factor was considered with two experimental variables. The first variable that we planned to investigate was the narrative about the source of the data provided to participants for memorization. Based on the conformity studies (Klucharev et al., 2009; Germar et al., 2014), we expected that the social narrative ("information which you memorize was provided by other people's responses") will reduce participants disagreement with the presented information and, in turn, will increase the HCD effect. At the same time, we expected much weaker learning in the *random* condition. In fact, the results of the second experiment (which plays the main part in supporting this conclusion) showed no difference in implicit learning in three conditions with different narratives of the provided information source. Interestingly, even when a narrative suggested absolute meaninglessness of the relationships between the photos and ratings (that ratings were random numbers generated by a computer) participants learned the hidden covariation.

The influence of the second variable that affected implicit learning was not directly envisioned before the first study. In Experiment 1, participants from the *social* and the *computer* conditions did not exhibit any implicit learning in contrast to the *random* condition. It appeared that the reason for this is not the nature of the provided narrative but the critical evaluation of the information during the learning phase, leading to explicit juxtaposition of the presented information and participants' opinion. In Experiment 1, participants in the first two conditions had to report any disagreement with the ratings while memorizing photographs with high ratings. Additionally, there were significantly more participants in these two conditions who reported that the attractiveness ratings in the learning phase were not plausible. The addition of the post-experimental agreement to the analysis revealed that it was the strongest factor of the learning effect. However, post-experimental agreement did not interact with the effect of learning in the second experiment, where there was no possibility to report disagreement during the learning phase in all three conditions. Perhaps, post-experimental disagreement reflects situational attitudes only when there was an actual possibility to actively report it during the learning phase. In the second experiment, post-experimental disagreement might reflect participants' traits rather than situational attitude to the presented material.

We considered three possible explanations for the disagreement effect: 1) dual-task situation, 2) absence of learning in concurrent disagreement condition, and 3) switch to analytical strategy which prevented implicit knowledge from manifestation in behavior. The first possibility was implausible from the theoretical and empirical points of view: most of the authors suggest that attention is critical for application of implicit knowledge and not for its acquisition (Frensch, Lin, & Buchner, 1998; Jiang & Leung, 2005; Waldron & Ashby, 2001). Two other explanations can be hardly distinguished with the designs like ours. However, we think that the explanation related to analytical strategy is preferable. First, it is well supported by the empirical data (Karpov & Moroshkina, 2014; Stamov Rossnagel, 2001; Lleras & Von Mühlenen, 2004; Ivanchei & Moroshkina, 2018). Second, it allows to fruitfully develop current models of implicit learning.

We suggest that decision-making strategies may be understood within dual-system approaches to implicit learning. There is a number of dual-system models (Sun, Slusarz, & Terry, 2005; Reber, 1993; Dienes, 2012), but let us consider one of the best developed models of implicit category learning – COVIS ("Competition between verbal and implicit system") by Ashby and colleagues (Ashby, Paul, & Maddox, 2011). This model assumes parallel work of two cognitive systems, which compete for the response. Which output is selected for the response is determined by the confidence of the two systems' decisions (based on the amount of evidence for one of the possible outcomes). The model also assumes that the outcome is initially biased towards the explicit system. Apart from these two factors, COVIS does not make any assumptions on the interaction between verbal and implicit systems. We suggest that the interaction between the systems is the very mechanism of decision-making strategy introduced above. Namely, the dominance of the implicit system corresponds to the holistic strategy, whereas the analytical strategy is the result of the verbal system dominance. COVIS predictions were usually tested with artificial stimuli and categories (see e.g., Waldron & Ashby, 2001). We suppose that in real-life situations, a person's attitude towards evaluated material can affect the contribution of different systems in behavior. Based on the studies demonstrating the impact of decision-making strategies on implicit learning (Karpov & Moroshkina, 2014; Stamov, Rossnagel, 2001; Lleras & Von Mühlenen, 2004; Ivanchei & Moroshkina, 2018), we suggest that critical attitude towards evaluated material increases the explicit system engagement in behavior. Facing some information which looks suspicious, e.g., coming from an unreliable source, a person starts to compare this information with her own relevant explicit memories. That is, the verbal memory system starts to dominate, i.e. the person switches to the analytical strategy. In this case, any implicit knowledge acquired before does not contribute to the final decisions.

Our results contribute to the discussion on the automaticity and control in implicit learning (Norman, Scott, Price, & Dienes, 2016; Ivanchei, 2014). Our study suggests that implicit learning in a social context may occur without any intention to learn and without knowledge of the presence of any regularity to be learned. The observable learning of hidden covariation contained in absolutely meaningless stimuli (randomly generated by a computer, according to the provided narrative) is striking. At the same time, the "critical attitude" findings suggest that although implicit learning occurs absolutely automatically, it may be prevented from showing itself in behavior as a result of participants' attitude towards the studied material. Thus, apart from conscious strategic control, there is a number of variables which may influence implicit knowledge manifestation in behavior. Future research should aim to clarify the relationships between the mechanisms underlying the effect of critical attitude and conscious control.

Our results can contribute to the investigation of social conformity in several ways. First, we demonstrated that socially relevant decisions can be biased not only by the reinforcement learning mechanisms in the brain, as concluded by Klucharev et al. (2009). In our study, attractiveness evaluation was shifted by mere observation of externally provided attractiveness ratings, without any actions and feedback which by definition are needed for reinforcement learning to occur. Moreover, when participants were able to report disagreement with the provided ratings (which is closer to reinforcement learning and to Klucharev et al. (2009) paradigm), no learning was observed.

Second, our study did not find any difference between learning in a social context and without it. This is inconsistent with Klucharev et al. (2009), who found stronger bias in a social version of the paradigm compared to computer-generated ratings. These distinctions may be explained by the involvement of different learning mechanisms: reinforcement learning and observational covariation learning which does not require actions/decisions and feedback (i.e. it is unsupervised). Independence from social context may mean that the latter form of learning is, in some sense, a more basic form of learning, relevant to processing any statistical information in the environment. Similar learning mechanisms for social perception were discussed by Dotsch, Hassin, and Todorov (2016), who observed shifts in facial evaluation by mere observation of different distributions of facial features (although with a much higher learning set of 500 faces). However, as noted above, there is a difference between observation of just positive examples of some category and learning of covariation with both positive and negative examples. The common feature of our results and the study by Dotsch, Hassin, and Todorov (2016) is the unsupervised nature of the learning. Our results are more comparable to the affective learning effect reported by Verosky and Todorov (2010). They found transfer of positive affect to facial morphs resembling those previously associated with positive behavior. This effect was preserved in the dual-task condition and with the direct instruction to ignore the facial resemblance (Verosky & Todorov, 2013), which suggests automatic nature of the effect. Our results are also congruent to Gamond et al. (2012), which reported HCD in the observational condition and failed to do so in the feedback condition. However, we think that Gamond et al.'s interpretation was wrong, and feedback learning by itself can underlie social influence. We argue that the critical attitude of participants was a confounding variable in Gamond's experiments. Namely, the feedback procedure forced participants to compare their opinions with the provided information.

Another finding that contributes to conformity studies is the demonstration that implicit learning influences not just overt responses (like in pioneering social pressure studies of Solomon Asch [Asch, 1951]) but also changes participants' internal criteria of attractiveness evaluation. This underlies transfer of learned covariation to new stimuli.

One interesting feature of our study is that it used stimuli derived from the same population as our participants (young adults of predominantly Caucasian appearance). Thus, participants might have well-formed criteria for evaluating attractiveness. Despite this, we observed HCD. We may suppose that for less frequently observed faces the learning effect may be even stronger. However, this question needs to be explored in separate studies.

One of the limitations of our study is the low amount of cases per cell. After removing aware participants, we had 13-14 participants in a counterbalancing subgroup. However, we applied mixed effects models which are more sensitive than traditional ANOVA, as every trial data is used for computing effects estimates.

Our study does not address the question of how long the effect of attractiveness bias lasts. Previous implicit learning research suggests that usual laboratory short-term learning sessions lose their effect rapidly if participants are exposed to new material that is not consistent with the hidden regularity (Mealor & Dienes, 2013; but see Hill et al., 1989, who reported self-perpetuation of HCD). Assuming that our participants meet loose and gathered-haired women after the experiment, they probably lose their learned hidden covariation sensitivity. However, we argue that if some regularity is present in the environment for a long time, implicit learning mechanisms described in this paper may have a strong influence on people's decisions in the social context. For example, the mechanisms observed in our study may be responsible for the adherence to the beauty standards imposed by fashion journals and the media in general. The media persistently presents models with a particular type of appearance and these are accompanied by implicitly assumed or explicitly stated high attractiveness ratings. According to the present study, if people do not deliberately compare their experience with the imposed information, they will change their minds in line with what society (or some of its influential members) wants.

Conclusions

Our study has demonstrated the possibility of change in facial attractiveness evaluation using a rather short implicit learning series. The learning did not depend on the narrative regarding learned material (the opinions of other people, automated (computer) measurements, or random numbers). What influenced implicit learning was the critical attitude towards the presented stimulus material. When participants were able to disagree with the presented ratings of the models during the learning phase, some of them (who in fact did not agree with these ratings) failed to demonstrate hidden covariation learning in the test phase. We believe this to be a consequence of the use of analytical decision-making strategy by these participants. Our study also contributes to conformity investigation by showing that basic information-processing algorithms can influence the evaluation of socially significant stimuli. Therefore, our research contributes to the studies of conformity and social perception as well as the investigation of automaticity and control in implicit learning.

Context

The main idea of the article occurred to the authors while discussing simple anchoring effect reported in Germar et al. (2014) study. The main theoretical idea on the decision-making strategy mediating social influence came from a series of authors' studies aimed to investigate the interplay between conscious and unconscious processing in implicit learning. The paradigm has been developed and used by the authors for the last several years. This study was a part of the project focused on the investigation of cognitive mechanisms involved in social processes.

Acknowledgements

This study has been supported by the Russian Foundation for Humanities (Project No. 15-36-01355). The authors are grateful to A. Karpov, who developed the paradigm used in the presented studies. The authors are also grateful to D. Shatalina and A. Izosimova for the assistance with the data collection. The authors are also grateful to the models who gave permission to use their photos for the demonstration of stimulus material in this paper.

References

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men; research in human relations* (pp. 177-190). Oxford, England: Carnegie Press.

Ashby, F. G., Paul, E. J., & Maddox, W. T. (2011). 4 COVIS. Formal approaches in categorization, 65.

Barker, L. A. & Andrade, J. (2006). Hidden covariation detection produces faster, not slower, social judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 636-641. https://doi.org/10.1037/0278-7393.32.3.636

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7), 1-23.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591-621.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. The journal of abnormal and social psychology, 51(3), 629.

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour*, *1*(1), 1. https://doi.org/10.1038/s41562-016-0001

Eskenazi, T., Montalan, B., Jacquot, A., Proust, J., Grèzes, J., & Conty, L. (2016). Social influence on metacognitive evaluations: The power of nonverbal cues. *The Quarterly Journal of Experimental Psychology*, *69*(11), 2233-2247.

Frensch, P. A., Lin, J., & Buchner, A. (1998). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research, 61*(2), 83–98. https://doi.org/10.1007/s004260050015

Gamond, L., Tallon-Baudry, C., Guyon, N., Lemaréchal, J. D., Hugueville, L., & George, N. (2012). Behavioral evidence for differences in social and non-social category learning. *Frontiers in psychology*, *3*, 291.

Germar, M., Schlemmer, A., Krug, K., Voss, A., & Mojzisch, A. (2014). Social influence and perceptual decision making: a diffusion model analysis. *Personality and Social Psychology Bulletin*, *40*(2), 217-231.

Hendrickx, H., De Houwer, J., Baeyens, F., Eelen, P., & Van Avermaet, E. (1997). Hidden covariation detection might be very hidden indeed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(1), 201-220. https://doi.org/10.1037/0278-7393.23.1.201

Hill, T., Lewicki, P., Czyzewska, M., & Boss, A. (1989). Self-perpetuating development of encoding biases in person perception. *Journal of Personality and Social Psychology, 57*(3), 373-387. https://doi.org/10.1037/0022-3514.57.3.373

Hill, T., Lewicki, P., Czyzewska, M., & Schuller, G. (1990). The role of learned inferential encoding rules in the perception of faces: Effects of nonconscious self-perpetuation of a bias. *Journal of Experimental Social Psychology, 26*(4), 350-371. https://doi.org/10.1037/0022-3514.57.3.373

Ivanchei, I. (2014). Theories of Implicit Learning: Contradictory Approaches to the Same Phenomenon or Consistent Descriptions of Different Types of Learning? *The Russian Journal of Cognitive Science, 1*(4), 4–30.

Ivanchei, I. I., & Moroshkina, N. V. (2018). The effect of subjective awareness measures on performance in artificial grammar learning task. *Consciousness and Cognition*, 57(January 2018), 116–133. https://doi.org/10.1016/j.concog.2017.11.010

Jiang, Y., & Leung, A. W. (2005). Implicit learning of ignored visual context. Psychonomic Bulletin & Review, 12(1), 100–6. https://doi.org/10.3758/BF03196353

Karpov A.D., Moroshkina N.V. (2014) The role of implicit learning in evaluating the psychological attributes of another person by their photoimage. *Litso chelovaka v nauke, iskusstve i praktike* [*The human face in science, art and practice].* pp. 93-106

Kinder, A., Shanks, D. R., Cock, J., & Tunney, R. J. (2003). Recollection, fluency, and the explicit/implicit distinction in artificial grammar learning. *Journal of Experimental Psychology. General*, *132*(4), 551–65. https://doi.org/10.1037/0096-3445.132.4.551

Kleisner, K., Chvátalová, V., & Flegr, J. (2014). Perceived intelligence is associated with measured intelligence in men but not women. *PloS One*, *9*(3), e81237. https://doi.org/10.1371/journal.pone.0081237

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement Learning Signal Predicts Social Conformity. *Neuron*, *61*(1), 140–151. https://doi.org/10.1016/j.neuron.2008.11.027

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).

Lewicki, P. (1986a). Nonconscious social information processing. New York: Academic Press

Lewicki P. (1986b). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(1)* 135-146. https://doi.org/10.1037/0278-7393.12.1.135

Lewicki, P., Hill, T., & Sasaki, I. (1989). Self-perpetuating development of encoding biases. *Journal of Experimental Psychology: General*, *118*(4), 323.

Lewicki, P., Hill, T., & Czyzewska, M. (1997). Hidden covariation detection: a fundamental and ubiquitous phenomenon. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 23*(1), 221–228. https://doi.org/10.1037/0278-7393.23.1.221

Lleras, A., & Von Mühlenen, A. (2004). Spatial context and top-down strategies in visual search. Spatial Vision, 17(4–5), 465–482. https://doi.org/10.1163/1568568041920113

Mealor, A. D., & Dienes, Z. (2013). Explicit feedback maintains implicit knowledge. Consciousness and Cognition, 22(3), 822–832. https://doi.org/10.1016/j.concog.2013.05.006

Moroshkina, N.V., & Karpov, A.D. (2015). The role of cognitive style of impulsivity-reflexivity in implicit learning (the example of the social perception tasks). *Eksperimental'naya psikhologiya [Experimental Psychology (Russia)]*, *8*(4), 61–76. https://doi.org/10.17759/exppsy.2015080405

Moroshkina N.V., Ivanchei I.I., Tikhonov R.V., Karpov A.D., Ovchinnikova I.V. (2018) Development and Approbation of The Russian Database of Neutral and Smiling Female Faces (RuNeS Faces). *Eksperimental'naya psikhologiya [Experimental Psychology (Russia)]*. 11(2), pp. 34–49. doi:10.17759/exppsy.2018110203. (In Russ., abstr. in Engl.)

Moroshkina N.V., Ivanchei I.I., Karpov A.D., Ovchinnikova I. (2019) The verbalization effect on implicit learning. In A. Cleeremans, V. Allakhverdov, M. Kuvaldina (Ed.), *Implicit Learning: 50 Years On* (pp. 189-207). London and New York: Routledge, Taylor and Francis group.

Norman, E., Scott, R. B., Price, M. C., & Dienes, Z. (2016). The relationship between strategic control and conscious structural knowledge in artificial grammar learning. Consciousness and Cognition, 42, 229–236. https://doi.org/10.1016/j.concog.2016.03.014

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods, 162*(1), 8-13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Price, M. C., & Norman, E. (2008). Intuitive decisions on the fringes of consciousness: Are they conscious and does it matter? *Judgment and Decision Making, 3*(1), 28–41.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, *6*(6), 855-863.

Reber, A. S. (1993). Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious (Oxford Psychology Series, No 19).

Stamov Roßnagel, C. (2001). Revealing hidden covariation detection: Evidence for implicit abstraction at study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1276.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychol. Rev.*, *112*(1), 159–92. doi:10.1037/0033-295X.112.1.159

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. doi:10.1126/science.185.4157.1124

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. Psychological Science, 21(6), 779-785.

Verosky, S. C., & Todorov, A. (2013). When physical similarity matters: Mechanisms underlying affective learning generalization to the evaluation of novel faces. *Journal of Experimental Social Psychology*, *49*(4), 661-669.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, 8, 168-176. https://doi.org/10.3758/BF03196154

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological science*, *17*(7), 592-598.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological science*, *22*(7), 894-900. https://doi.org/10.1177/0956797611411057

Zebrowitz, L. A. (1996). Physical appearance as a basis of stereotyping. *Stereotypes and stereotyping*, 79-120.

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate Social Perception at Zero Acquaintance: The Affordances of a Gibsonian Approach. *Personality and Social Psychology Review, 1*(3), 204–223. https://doi.org/10.1207/s15327957pspr0103_2

Zizak, D. M., & Reber, A. S. (2004). Implicit preferences: the role(s) of familiarity in the structural mere exposure effect. *Consciousness and Cognition, 13*(2), 336–62. https://doi.org/10.1016/j.concog.2003.12.003

Appendix

The questions used in the post-experimental interview:

Q1. “Please indicate which particular appearance features you associated with high or low attractiveness”.

Q2. “Did the attractiveness ratings seem plausible to you while you were memorizing the women’s faces during the first stage?” (Possible answers: “Yes, the ratings were congruent with my own impressions”, “No, the ratings seemed inadequate, I didn’t agree with most of them”, “I had not thought of it until you asked me about it”).

Q3. “Did you notice that during the first stage the models with high ratings differed from the models with low ratings by a common characteristic? If so, what was it?”.

Q4. “Please indicate whether you rely on the opinion of other people when you assess the attractiveness of a person you see for the first time”. (Possible answers: “Yes, other people’s opinions are important to me”, “Yes, but I rely only on the opinions of one or two people close to me”, “I do not care what other people think, I'm only guided by my own impressions”).