Acta Psychologica 151 (2014) 206–213

Contents lists available at ScienceDirect

Acta Psychologica

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

How to tell a wife from a hat: Affective feedback in perceptual categorization

Andrey Chetverikov [*], Margarita Filippova [1]

Department of Psychology, Saint Petersburg State University, 6 Makarova emb., Saint Petersburg 191002, Russia

ARTICLE INFO

Article history:
Received 9 February 2014
Received in revised form 22 June 2014
Accepted 25 June 2014
Available online 19 July 2014

PsycINFO codes:
2300
2323
2360

Keywords:
Perceptual hypotheses
Categorization
Gestalt
Ambiguity
Errors
Affective feedback

ABSTRACT

How do people understand that their perception is correct? In line with the recurring idea of perception as prediction, the affective feedback account of hypotheses testing suggests that correct perceptual predictions are reinforced with positive affect. In four experiments, we tested whether correct categorization of a degraded image will lead to more positive liking ratings. The obtained findings supported the proposed approach: subjects liked the images they were able to perceive correctly more than others. Importantly, these findings were independent of the initial affective valence of stimuli. A further investigation demonstrated that this effect exists only when answers are at least moderately confident. The obtained findings add to the growing amount of literature on the role of affect in basic cognitive processing.

1. Introduction

Usually, people do not question the veridicality of their perception. However, there are situations when observers may doubt their senses. In milder cases, people can be uncertain that what is seen is seen correctly — as, for example, when someone is unsure that the face in the crowd is indeed the face of a friend. In worse cases, people may constantly check the validity of their memories or perceptions. Such distortions have been described as part of "pathological worry" in studies of general anxiety disorder or of "pathological doubt" in the case of obsessive–compulsive disorder (Starcevic & Berle, 2006; Tolin et al., 2001). In contrast, sometimes people do not question the veridicality of their perceptions even when they are clearly incorrect. An example is the famous case of the man who mistook his wife for a hat (Sacks, 2011). Yet, how do we know, that we have perceived something correctly? To answer this question, it is necessary to look into the mechanisms of perception.

Perception is not a passive process. Von Helmholtz (1866) suggested that perception is guided by unconscious inferences, Bruner (1957) and Gregory (1997) used the notion of perceptual hypotheses, and currently this idea recurs in predictive coding models (Friston, 2010; Hohwy, 2012, 2013). A recent proposal is that at each level of processing affective feedback reinforces the development of a realistic model of the world (Allakhverdov & Gershkovich, 2010; Chetverikov, 2014; Chetverikov, Jóhannesson, & Kristjánsson, 2014). This approach, coined as the affective feedback account of hypotheses testing, suggests that at each level of processing our cognitive system tries to predict what our environment is. If these predictions are correct, then we are reinforced with positive affect. If they are not, then we experience negative affect that facilitates the changes of hypotheses. Hence the experience of veridical perception is different from the experience of perceptual errors in its affective valence, allowing people to distinguish between the two.

Despite the intuitive appeal of the affective feedback idea, it lacks empirical testing. Although many findings are in favor of the proposed approach, they are mostly indirect. For example, the effects of processing fluency on preferences (Bornstein & D'Agostino, 1994; Reber, Schwarz, & Winkielman, 2004) indicate that items processed with more ease are rated as more pleasant than the rest. However, processing fluency is a natural consequence of our expectations. To give an example, in Experiment 1 of Reber, Winkielman, and Schwarz (1998) processing fluency was manipulated by presenting a matching or

* Corresponding author at: Department of Psychology, Saint Petersburg State University, Room 7, Department of Psychology, Saint Petersburg State University, nab. Makarova 6, Saint Petersburg 191002, Russia. Tel.: +7 9110902215.
E-mail addresses: a.chetverikov@psy.spbu.ru (A. Chetverikov), margarita.g.filippova@gmail.com (M. Filippova).
[1] Tel.: +7 9213312880.

http://dx.doi.org/10.1016/j.actpsy.2014.06.012
0001-6918/© 2014 Elsevier B.V. All rights reserved.

non-matching prime before the picture. Observers' predictions were confirmed by the following stimuli in the case of matching prime or contradicted by non-matching one. This led both to a decrease in reaction times and to positive affect. Other cases of fluency manipulations, such as increased contrast or distortions of symmetry can be treated as manipulations of prediction accuracy, because decreased uncertainty helps to provide correct predictions (see also Van de Cruys & Wagemans, 2011). Yet, this and similar evidence are mostly circumstantial for the proposed approach.

Also in favor of the proposed approach, Chetverikov (2014) and Chetverikov et al. (2014) demonstrated that even in the absence of external feedback, errors in recognition and visual search with brief displays result in a decreased preference ratings. Their reasoning was that errors could be interpreted as inconsistent prediction. Consequently, they should be followed by negative affective feedback. However, it is unclear, whether a simple act of perception can be treated similarly to more complex tasks, such as the ones used in these studies.

More direct evidence comes from the study by Muth and Carbon (2013). The authors investigated the "aha" experience associated with the perception of a hardly-detectable Mooney faces on ambiguous background. The observers repeatedly judged the attractiveness of images, some of which contained the Mooney faces while others did not. When observers finally found the face, the ratings were more positive than before. According to the affective feedback account, when observers were able to make correct perceptual hypotheses, they received positive affective feedback. Whether this effect will generalize to stimuli other than faces is unknown, however.

To sum up, there is evidence in favor of the general idea of affective feedback in hypotheses testing. The effects of making a correct perceptual prediction on preferences have, however, only been measured in one study (Muth & Carbon, 2013). The aim of the present study was to provide further evidence that accurate hypotheses about the content of perceived images evoke positive affect.

In four experiments reported here we used categorization task with ambiguous images to test, whether subjects able to perceive the object in these images will like them more than those who do not. Unlike Muth and Carbon (2013), we held exposure time constant for all stimuli and controlled for the effect of processing fluency by incorporating response time into the analyses. Experiment 1 provided initial data on the effects of correct perception on preferences. Experiment 2 demonstrated that this effect could not be attributed to initial affective valence of the stimuli. Experiment 3 further demonstrated that subjects like correctly categorized stimuli more than incorrectly categorized ones only after at least moderately confident answers. Finally, Experiment 4 replicated the findings of Experiment 2 and Experiment 3.

2. Experiment 1

In Experiment 1 we tested the hypothesis that perception of an object will evoke positive affect.

2.1. Method

2.1.1. Participants

Twenty undergraduate psychology students (17 females, age $Mdn = 20$) at Saint Petersburg State University voluntarily participated. No incentive was provided for taking part.

2.1.2. Materials and procedure

A set of 28 "hidden figure" black-and-white images, similar to the famous Dalmatian picture, were used as stimuli. These images depicted humans (N = 9), animals (N = 10), and inanimate objects (N = 10). Participants were informed that they would be participating in a study of perception, and that their task was to categorize images using the three aforementioned categories. Some of the images contained both humans and objects ("man sitting on a bench"). Subjects were instructed that if they saw both objects and human or object and animal they should categorize it as human or animal, respectively. The categorization task allows testing the accuracy of perception and lacks the ambiguity of free report interpretation (see the General discussion).

There were 3 training trials and 25 test trials. The trial sequence is presented in Fig. 1. After being exposed to a stimulus for 1000 ms, participants categorized it using the keyboard arrow keys ("left" — human, "down" — animal, "right" — object). Participants were then asked to rate each stimulus for liking ("How much do you like the presented image?") using a 100-point rating scale. No feedback about the accuracy of categorization was provided.

The first three images, one for each of the categories, were presented in the same order for each participant. Participants were repeatedly exposed to these images until they categorized them correctly. The remaining images were presented only once. The order in which the remaining images were presented was randomized.

2.2. Results

2.2.1. Categorization

The average categorization accuracy was well above chance, $M = 0.78$ [0.66, 0.87],[2] $t(59) = 22.13$, $p < .001$.

2.2.2. Liking

The liking ratings were analyzed using linear mixed-effects regression, LMER, with the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2013). In contrast to a more traditional approach with data aggregation and repeated-measures ANOVA analysis, LMER allows controlling for the variance associated with random factors without data aggregation (see Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012).[3] By using random effects for subjects and stimuli, we controlled for the influence of different mean ratings associated with these variables. For the sake of brevity, we present only the *F* tests from the LMER results here (type III Wald *F* tests with Kenward–Roger degrees of freedom approximation).

Average liking ratings and their confidence intervals are presented in Fig. 2A. Ratings were more positive after correct answers than after errors, $M = 54.28$ [51.88, 56.68] vs. $M = 44.57$ [40.49, 48.66], $F(1, 469) = 11.86$, $p < .001$. To assess the effect of fluency of processing, we repeated the analysis, this time including response time as predictor. Response time was logarithmically transformed to reduce the influence of extreme values (Fazio, 1990). If subjects' ratings were more positive because some stimuli were processed more fluently than others were, then there should be a negative effect of response time on liking. Indeed, we found a significant negative effect of response time, $F(1, 486) = 11.89$, $p < .001$. However, the effect of answer correctness still was significant, $F(1, 468) = 7.73$, $p = .006$, indicating that differences between correct and incorrect answers cannot be fully explained by differences in processing fluency.

We then analyzed stimuli by answer category to see if the attribution of stimuli to specific categories may explain the effects obtained for categorization accuracy. Table 1 shows means and confidence intervals for liking split by answer category. A two-way LMER with answer category and answer correctness showed significant effect of answer correctness, $F(1, 455) = 9.01$, $p = .003$, and a main effect of answer category, $F(2, 389) = 9.03$, $p < .001$. The interaction effect was not significant, $F(2, 118) = 1.08$, $p = .342$. Subjects rated images categorized as humans, $t(275) = 1.79$, $p = .074$, and animals, $t(138) = 2.36$, $p = .020$, as more likeable when the categorization was correct. For stimuli

[2] Here and in what follows, we present 95% confidence intervals in square brackets after mean values.

[3] The same analyses repeated with by-subject aggregation and repeated-measures ANOVA yielded the same results in regard to the decisions about effects' statistical significance. Confidence intervals were wider in the case of ANOVA than in the case of LMER as expected due to the data aggregation.

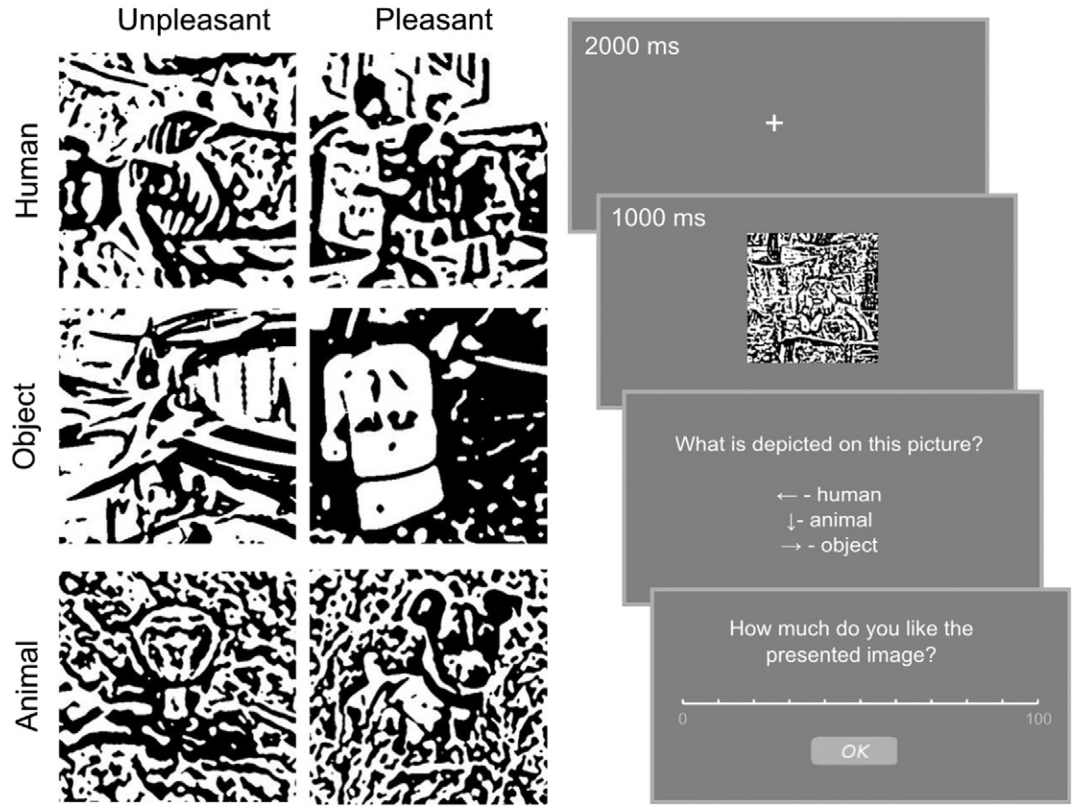

Fig. 1. Examples of stimuli (left) and trial sequence (right). In Experiments 1 and 3 neutral images were used (not shown here due to the possible copyright issues).

categorized as objects the effect of answer correctness was positive but not significant, $t(97) = 0.48$, $p = .635$.

2.3. Discussion

The results of Experiment 1 support our hypothesis by demonstrating that correct categorization of stimuli leads to more positive liking ratings. This effect is unlikely to be a result of increased fluency of processing as exposure time was kept constant and even with the addition of response time to the regression model the effect of accuracy remained significant. Moreover, we controlled for the influence of differences in stimuli and subjects. It is therefore doubtful that some of the within-stimuli variables (such as the difficulty of categorization) or within-subject variable (such as mood) can explain the observed results.

The Experiment 1 provides preliminary evidence that when subjects were able to perceive the figure hidden in the stimuli, they experienced positive affect. We used a forced-choice categorization procedure and it is possible that some of the answers were correct purely by chance. Thus, only some of the correct trials created the “Aha!” moment and the obtained difference in liking ratings is probably lower than the real effect.

However, there might also be an alternative explanation for the obtained findings. Our stimuli were generally positive or mildly neutral.

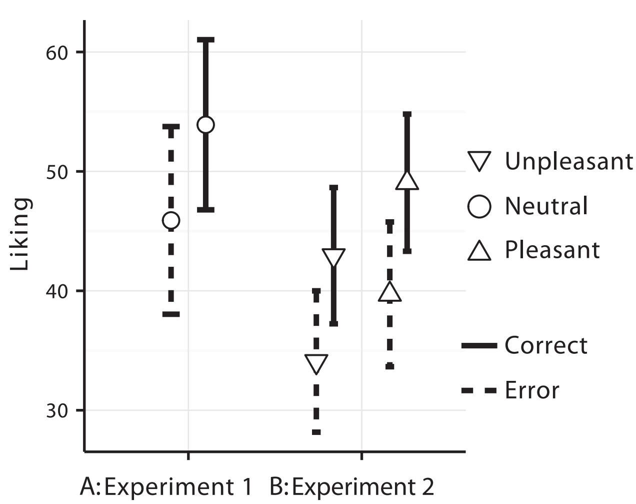

Fig. 2. Liking ratings as a function of categorization accuracy and affective valence of the stimuli (Experiments 1 and 2). *Note.* Bars represent 95% confidence intervals.

Table 1
Means (and standard deviations) of ratings as function of answer category, answer correctness, and experimental conditions.

		Exp. 1	Exp. 2		Exp. 3	
		Neutral	Pleasant	Unpleasant	Liking	Confidence
Animal	Correct	58.47	53.97	44.66	55.93	69.37
		(23.39)	(25.34)	(25.10)	(26.97)	(28.79)
	Error	45.07	42.21	33.01	40.56	42.45
		(23.28)	(24.61)	(19.85)	(25.44)	(28.79)
Human	Correct	59.73	58.79	47.71	58.25	76.82
		(24.41)	(25.74)	(25.74)	(27.27)	(26.58)
	Error	51.27	36.98	36.27	41.36	42.10
		(20.37)	(22.52)	(21.85)	(18.74)	(25.54)
Object	Correct	45.21	40.50	39.98	45.57	61.08
		(22.01)	(23.71)	(23.29)	(25.85)	(29.94)
	Error	42.28	27.30	27.15	35.99	46.89
		(20.56)	(19.87)	(18.60)	(23.84)	(33.79)

Note. Data in each column, except the confidence column of Experiment 3, represent liking ratings on a 100-point rating scale. The confidence column represents confidence ratings on a 100-point scale.

It is possible that subjects liked them more simply because what they saw was likable compared to a chaotic pattern of black and white spots. The aim of Experiment 2 was to test this explanation by using both positive and negative stimuli.

3. Experiment 2

Experiment 2 was designed to make sure that the difference between correct and incorrect answers observed in Experiment 1 was not due to the initial affective valence of stimuli. It is possible that if subjects understood what is depicted on the picture, then they give it a more positive rating simply because the content of the picture was positive. This way of reasoning suggests that subjects will rate unpleasant pictures more negatively when they are able to determine the content of these pictures. To test this hypothesis, in Experiment 2 we used both pleasant and unpleasant pictures.

3.1. Method

3.1.1. Participants

Twenty-three psychology students (15 females, age $Mdn = 22$) at Saint Petersburg State University voluntarily participated. No incentive was provided for taking part. None of the subjects participated in Experiment 1.

3.1.2. Materials and procedure

Stimuli included 54 pleasant and 60 unpleasant images. The images were created from pleasant and unpleasant images, which were either taken from GAPED database (Dan-Glauser & Scherer, 2011) or found in Internet. Pleasant stimuli also included the images from the previous experiment. The unpleasant images were comprised from different categories, including snakes, spiders, aggressive animals, rotten food, broken things, suffering people, etc. Pleasant stimuli included cute animals, smiling people, nice looking objects, etc. The procedure was the same as in Experiment 1.

3.2. Results

3.2.1. Categorization

We used the same data analysis approach as in Experiment 1. The average categorization accuracy was lower than in Experiment 1, but still above chance level, $M = 0.68$ [0.60, 0.75], $t(137) = 22.43$, $p < .001$. A comparison of unpleasant and pleasant images indicated that unpleasant images were more difficult for subjects than pleasant ones, $M = 0.65$ [0.61, 0.69] vs. $M = 0.71$ [0.67, 0.75], $t(135) = -2.00$, $p = .048$.

3.2.2. Liking

A two-way LMER with answer correctness and pleasantness revealed that subjects liked pleasant stimuli more than unpleasant, $F(1, 119) = 9.61$, $p = .002$, and correctly categorized stimuli more than incorrectly categorized, $F(1, 2540) = 80.69$, $p < .001$. The interaction effect was not significant, $F(1, 2545) = 0.06$, $p = .809$ (see Fig. 2B). Unpleasant images were liked less than pleasant ones, when subjects correctly identified the image category, $M = 44.46$ [42.91, 46.20] vs. $M = 50.62$ [48.99, 52.28], $t(133) = 3.16$, $p = .002$. Interestingly, when subjects made an error, they also liked unpleasant images less than pleasant ones, $M = 30.86$ [28.94, 32.84] vs. $M = 34.27$ [31.73, 36.75], $t(257) = 2.41$, $p = .017$. After correct answers liking ratings were higher both for unpleasant images, $M = 30.86$ [29.06, 32.62] vs. $M = 44.46$ [42.96, 45.92], $t(2465) = 6.39$, $p < .001$, and for pleasant ones, $M = 34.27$ [32.07, 36.57] vs. $M = 50.62$ [49.07, 52.30], $t(2465) = 6.39$, $p < .001$.

As in the Experiment 1, we also analyzed the effects on response time and answer category. We found a significant negative effect of response time, $F(1, 2654) = 32.53$, $p < .001$. However, the effect of answer correctness was also significant, $F(1, 2549) = 71.67$, $p < .001$, again showing that the difference in liking between correct and incorrect answers cannot be fully explained by differences in processing fluency.

An analysis of stimuli split by answer category and pleasantness indicated that in each case correct categorization was associated with higher ratings (see Table 1). A LMER model including the three-way interaction of answer category, answer correctness and image pleasantness indicated a significant effect of all three main factors: $F(2, 1767) = 48.54$, $p < .001$ for category, $F(1, 2229) = 83.43$, $p < .001$ for correctness, and $F(1, 119) = 10.06$, $p = .002$ for pleasantness. Other effects were not significant.

3.3. Discussion

The results of Experiment 2 demonstrate that the effect of answer accuracy on liking ratings cannot be attributed to the initially positive ratings of categorized stimuli. Subjects rated both pleasant and unpleasant items as more likable when they correctly determined the category of image. Thus, although the positive appraisal of the content of perception increases liking ratings, the act of correct perceiving itself makes an independent contribution.

It might seem strange that when observers correctly recognize the unpleasant images they like it more than when they do not perceive it. This is explained by the fact that much of the negative evaluation of typical stimuli is related to the perceptual features, such as sharp angles or irregular patterns. In addition, it is possible that degraded image is less likely to evoke negative memory associations. While the accurate categorization is pleasant, the image categorized or its associations might be unpleasant, leading to negative ratings of correctly identified unpleasant objects in real life.

One of the potential drawbacks of the present study is that even for pleasant stimuli mean ratings were not very positive, and even lower than in the Experiment 1. However, a comparison of data for different categories of stimuli presented in Table 1 shows that for correctly identified pleasant stimuli the ratings in Experiment 1 and Experiment 2 were similar. Therefore, the lower mean ratings are explained by lower accuracy of subjects in the present experiment. In addition, it is unreasonable to expect very high positive liking ratings from black-and-white degraded stimuli like the ones that were used in this experiment.

4. Experiment 3

Experiment 1 also demonstrated that the effect of answer correctness on liking was significant only for humans and animals. Ratings of stimuli categorized as objects stimuli were the same, whether objects were categorized correctly or not. This could reflect the initial valence of stimuli as well, but it could also be explained by the fact that observers were less confident about answers in this category. Our instruction stated that images containing both human and object or animal and object should be categorized as human or animal, respectively. If subjects saw an object they might feel unsure that they did not miss something else. In order to test, whether the effect of perception accuracy on liking depends on confidence, Experiment 3 measured both confidence and liking.

Confidence judgments reflect the amount of evidence in favor of the chosen alternative (e.g., Koriat, 2012; Ratcliff & Starns, 2013). The internal feedback about the accuracy of hypothesis, such as the one proposed by the affective feedback account of hypotheses testing, requires the information about the correctness of hypotheses to be acquired — otherwise the feedback lacks basis. Low-confidence answers mean, on average, that little or no information was acquired. We expected that low-confidence answers, whether correct or incorrect, would be associated with similar liking ratings. For more confident answers, on the other hand, we expected a difference between correct and incorrect answers in liking. Consequently, if "Object" category has lower confidence than the others, then the effect of answer correctness on liking should also be weaker.

4.1. Method

4.1.1. Participants

Forty undergraduate psychology students (31 females, age $Mdn = 19$) at Saint Petersburg State University voluntarily participated. No incentive was provided for taking part. None of the subjects participated in the previous experiments.

4.1.2. Materials and procedure

The procedure and materials were identical to Experiment 1, with the exception that after providing their answers about the image category, participants made confidence ratings in addition to liking ratings. Participants were asked to rate their confidence in the accuracy of their categorization answer using a 100-point rating scale. To reduce the chances of participants' answers to the liking question influencing their answers to the confidence question, the second scale appeared only after participants clicked the "OK" button that was positioned below the middle point of the scale. To eliminate order effects, the order of rating scales was counterbalanced.

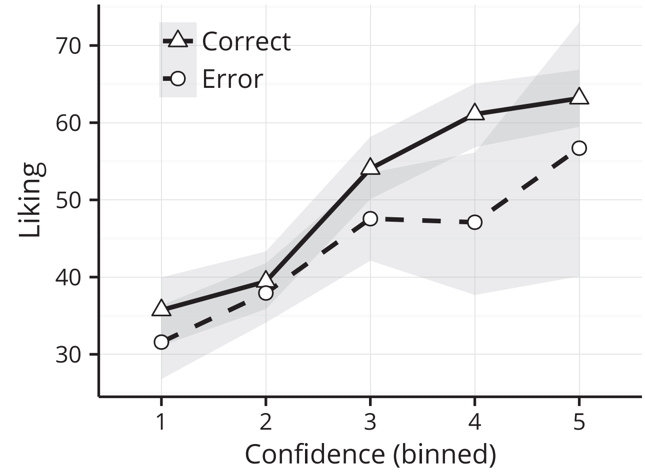

Fig. 3. Difference between correct and incorrect answers as a function of confidence in Experiment 3. *Note.* Confidence ratings were normalized and split into five equally sized bins, so that each bin contains correct answers and errors with similar confidence. Shaded areas represent 95% confidence intervals.

4.2. Results

4.2.1. Categorization

We used the same data analysis approach as in Experiment 1. The average categorization accuracy was above chance, $M = 0.74$ [0.66, 0.81], $t(119) = 23.87$, $p < .001$.

4.2.2. Liking and confidence ratings

Analyses of the effects of accuracy on liking yielded the same results as in previous experiments. For correct answers the liking ratings were higher than for erroneous ones, $M = 53.13$ [51.16, 55.10] vs. $M = 38.97$ [36.07, 41.87], $F(1, 951) = 14.79$, $p < .001$. Similarly, correct answers were more confident than the wrong ones, $M = 68.79$ [66.68, 70.91] vs. $M = 44.06$ [40.39, 47.73], $F(1, 966) = 69.13$, $p < .001$. Response time significantly affected both liking ratings, $F(1, 981) = 34.41$, $p < .001$, and confidence ratings, $F(1, 975) = 120.70$, $p < .001$, but in both cases the effect of accuracy remained significant, $F(1, 949) = 7.69$, $p = .006$ and $F(1, 964) = 43.60$, $p < .001$, respectively.

The analysis of liking ratings split by answer category also replicated the results of Experiment 1. Two-way LMER with answer correctness and answer category on liking showed a significant main effect of answer correctness, $F(1, 946) = 13.88$, $p < .001$, and a significant effect of answer category, $F(2, 694) = 4.70$, $p = .009$. Interestingly, analyses of confidence ratings also demonstrated not only an influence of answer correctness, $F(1, 961) = 57.76$, $p < .001$, but also of answer category, $F(2, 813) = 4.74$, $p = .009$.

As shown in Table 1, after correct answers confidence and liking ratings followed the same pattern: ratings for "human" answers were higher than ratings for "animal" answers, which, in turn, were higher than ratings for "object" answers. After errors, on the other hand, ratings were similar for "human" and "animal" categories, and for "object" category ratings were lower in the case of liking, but higher in the case of confidence. It is important to note that the difference in confidence ratings between correct answers and errors for "object" category was smaller than for two other categories. This supports the idea that the lower effect of accuracy in this category may be explained by confidence.

4.2.3. Liking ratings split by levels of confidence

We then turned to the main question of the present experiment. Does the difference between liking ratings after correct and incorrect categorization depends on confidence? To answer this question, it is necessary to have comparable estimates of confidence for each subject. Thus, we normalized confidence ratings (Z-transformation was applied separately for each subject). Transformed confidence ratings were split in five equally sized bins. The number of bins was set to five to provide detailed estimates of liking ratings while keeping at least ten errors in each bin. The resulting liking ratings for each bin are shown in Fig. 3. Each bin includes both correct and incorrect classifications with the confidence ratings in the same range. We then used LMER with preset contrasts to compare liking after correct and incorrect answers on each level of confidence. The interaction between confidence bin and answer correctness was only marginally significant, $F(1, 963) = 3.33$, $p = .068$. Contrasts testing indicated that the effect of answer correctness was nearly zero for the first two levels of confidence, $t(948) = 0.07$, $p = .943$ and $t(942) = 0.05$, $p = .964$, then increased in the third bin, $t(942) = 1.53$, $p = .126$, and reached significance in the fourth bin, $t(939) = 2.26$, $p = .024$. Finally, with the highest confidence level the effect of correctness was not significant again, $t(946) = 0.53$, $p = .596$.

4.3. Results

The results of Experiment 3 demonstrate that correct answers led to higher liking ratings only when subjects were moderately confident in their answers. Very low or very high confidence was associated with non-significant differences in liking after correct and incorrect answers. As predicted, we found the effect of correctness on liking only when subjects had some information regarding the stimuli and not when they simply were able to guess correctly. However, although the paired comparisons indicated a predicted pattern, the interaction effect that supports the conclusions was only marginally significant. We decided to run a replication study with the set of stimuli used in Experiment 2.

5. Experiment 4

This study was an exact replication of Experiment 3, with the exception that the stimuli used were taken from Experiment 2. We expected to obtain the effects observed in Experiments 2 and 3: the effect of answer accuracy on liking should be observed both for pleasant and for unpleasant stimuli, and it should be more pronounced at higher levels of confidence.

5.1. Method

5.1.1. Participants

Thirty psychology students (24 females, age $Mdn = 21$) at Saint Petersburg State University voluntarily participated. No incentive was provided for taking part. None of the subjects participated in the previous experiments. Two subjects were excluded from the analysis because they used the same liking rating for more than 95% of the stimuli.

5.1.2. Materials and procedure

The procedure was identical to Experiment 3, but the stimuli used were from Experiment 2 (a set of 50 pleasant and 64 unpleasant images).

5.2. Results

5.2.1. Categorization

The average categorization accuracy was above chance, $M = 0.78$ [0.69, 0.84], $t(55) = 46.44$, $p < .001$. A comparison of unpleasant and pleasant images indicated that unpleasant images were no more difficult for subjects than pleasant ones, $M = 0.71$ [0.69, 0.73] vs. $M = 0.74$ [0.71, 0.76], $t(53) = -1.61$, $p = .113$.

5.2.2. Liking and confidence ratings

Subjects liked pleasant stimuli more than unpleasant, $F(1, 123) = 9.97$, $p = .002$, and correctly categorized stimuli more than incorrectly categorized, $F(1, 3033) = 50.33$, $p < .001$. Unlike Experiment 2, the interaction effect of answer accuracy and image pleasantness was significant, $F(1, 3039) = 5.62$, $p = .018$. The significant interaction term was due to the fact that the difference between pleasant and unpleasant images was significant when subjects correctly determined the category of image, $t(130.1) = 4.39$, $p < .001$, but not when they made an error, $t(273) = 1.48$, $p = .139$. Correct answers led to higher liking both for unpleasant images, $t(2993.1) = 3.49$, $p < .001$, and for pleasant ones, $t(3070.2) = 6.46$, $p < .001$.

Correct answers were more confident than errors, $M = 75.26$ [74.12, 76.41] vs. $M = 44.06$ [40.39, 47.73], $F(1, 966) = 69.13$, $p < .001$. Pleasantness of image had no significant effect on confidence, $M = 69.83$ [68.30, 71.35] vs. $M = 65.89$ [64.40, 67.38], $F(1, 115) = 1.61$, $p = .207$.

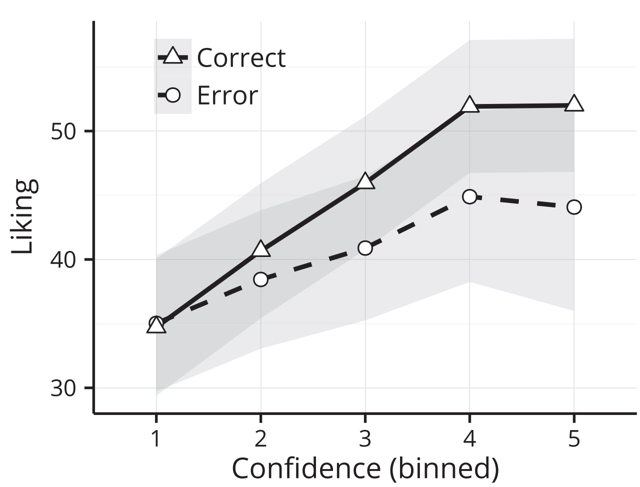

Fig. 4. Difference between correct and incorrect answers as a function of confidence in Experiment 4. *Note*. Confidence ratings were normalized and split into five equally sized bins, so that each bin contains correct answers and errors with similar confidence. Shaded areas represent 95% confidence intervals.

5.2.3. Liking ratings split by level of confidence

Confidence ratings were standardized and split into bins, following the same routine as in Experiment 3. There was a significant interaction effect of confidence bin and answer correctness, $F(1, 3208) = 6.54$, $p = .003$ and $t(3236) = 2.87$, $p = .004$. Moreover, it was also significant when the confidence was at the highest level, $t(3229) = 2.31$, $p = .021$.

5.3. Discussion

Experiment 4 provided a replication of the results obtained in Experiment 2: both for pleasant and for unpleasant images the correct answers were followed by more positive liking ratings than errors. Most importantly, we also replicated the results of Experiment 3. Only when subjects were confident in their answers we found a difference between correct answers and errors in liking ratings. Even the most confident errors turned out to be followed by less positive liking ratings than the most confident correct answers.

6. General discussion

6.1. Liking as a function of accuracy and confidence

Four experiments reported here demonstrate that when subjects were able to categorize the presented stimuli correctly, they liked these stimuli more than when they were not able to do it. According to the proposed approach this indicates that accurate perceptual hypotheses evoke positive affect. That is, when perceiver's cognitive system can generate an accurate prediction about the content of an image, separating it into the figure and ground, positive affect reinforces this prediction.

The obtained results cannot be explained by the affective valence of stimuli, because both pleasant and unpleasant stimuli in Experiments 2 and 4 demonstrated the effect of answer accuracy on liking ratings. Thus affective valence of stimuli is not important for the observed effect. They also cannot be explained by the attribution of stimuli to specific categories, as the effect of correctness was positive for all answer categories. However, the amount of acquired information in support of the chosen hypothesis measured by confidence ratings does seem to be important. As discussed above only at least moderately confident answers exhibit the accuracy-dependent influence on liking.

One might argue that if observers have some confidence in correct and incorrect answers, then the feedback about the accuracy of hypotheses, and hence the liking ratings, should be the same. However this argument is correct only if the mechanisms involved in confidence estimation and liking are the same, which may not be the case. Although there are different models of processes leading to confidence judgments (e.g., Koriat, 2012; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2013), the relationship between perceptual decisions and liking is yet to be studied.

The unexpected finding from Experiment 3 is that answers that observers are highly confident about also did not have an influence on liking ratings. Most probably, highly confident errors were indistinguishable from highly confident correct answers. This finding also warrants further investigation. However, this result was not replicated in Experiment 4 and should be treated with caution.

The obtained findings demonstrate that the effects of accuracy previously demonstrated for recognition task (Chetverikov, 2014) and for visual search (Chetverikov et al., 2014) can be also found in categorization. This suggests that the observed changes in liking may be governed by a general mechanism supporting the claims of the proposed approach. To reiterate, the affective feedback account of hypotheses testing suggests that correct hypotheses are reinforced with positive affect even in the absence of external feedback. That is, if we predict that the object we see is an apple and it is an apple, then we will like this object more, than if we predict that it is a tomato. Thus the obtained findings support theories of error-monitoring that postulate the involvement of affect in this process (Hajcak, McDonald, & Simons, 2004; Luu, Collins, & Tucker, 2000; Wiswede, Münte, Goschke, & Rüsseler, 2009). This affect then spreads to the stimuli at hand. To be clear, this explanation does not suggest that affect is initially bound to the motor response (that is, the action of pressing the button). We assume that it is the hypothesis – here, the classification decision, – that evokes the affect. However, in the present study we did not separate the motor and the cognitive component and the provided explanation needs further testing.

6.2. Alternative interpretations

There are several possible alternative explanations for the current results. First, it is possible that the liking ratings reflected fluency of stimuli processing and not affective feedback. However, this explanation is undermined by the fact that exposure time was kept constant, we included random effects for stimuli, and even after the addition of response time to the regression equation the effect of answer correctness was still significant. In addition, as we discussed in the Introduction, some of the findings on processing fluency studies can be explained by the affective feedback account of hypothesis testing. Still, the theoretical lack of clarity regarding the measurements of fluency (see, for example, Box 1 on p. 238 in Oppenheimer (2008) and the discussion on p. 24 in Chetverikov (2014)) makes it difficult to refute this explanation.

Second, we asked subjects to fulfill a categorization task, and it is possible that in the absence of such task we would not observe the effect of perceiving an object on liking. Perhaps, this task presents an interesting puzzle for subjects, and they feel good when they are able to solve it correctly. However, as argued by Bruner (1957) any perception involves categorization. Our task only directs this process in a manner that is convenient for measurement. If we would not suggest some categories for subjects, they will use their own. Thus although this question need further study, the categorization task is unlikely to be responsible for the observed effect.

For the same reasons it is hard to disentangle positive affect elicited by correct and negative affect elicited by errors using the perceptual categorization task. Basing on the results of the previous studies with recognition and visual search (Chetverikov, 2014; Chetverikov et al., 2014), we expect that effect of correct and erroneous perception can be separated by varying the amount of information that supports the correct interpretation. Yet currently this issue remains unresolved.

Third, it is possible to explain the obtained findings as a result of differences in confidence. That is, people like stimuli after more confident answers more than after less confident answers. However, Experiments 3 and 4 demonstrated that even when correct answers and errors had similar confidence liking ratings were still more positive after correct answers. It is also not clear why confidence would increase the liking rating, unless there is an affective component in confidence itself. If it is not pleasant to be confident, then why we would like the stimuli after confident answers more than after less confident ones? On the other hand, if it is pleasant to be confident, then both confidence and liking ratings measure affect to some degree. We would not argue with that statement. In fact, our recent study indicates that irrelevant pleasant or unpleasant stimuli can influence confidence ratings (Chetverikov, 2013), directly supporting the idea that feeling of confidence has an affective component. We suggest that affect is evoked by the hypotheses-testing process and the target stimulus is only a convenient object for the attribution of that affect. One's own choice is even more convenient object and confidence ratings may provide a convenient way to measure the affect attributed to it. Accordingly, it is possible that this attribution can be guided through traditional attribution manipulation (Schwarz & Clore, 1983), although the present study did not test it.

The proposed account is corroborated by the recent data that links stimuli and response conflict with aversive tendencies (Aarts, De Houwer, & Pourtois, 2012, 2013; Chetverikov & Kristjánsson, 2014; Dreisbach & Fischer, 2012; Fritz & Dreisbach, 2013; Martiny-Huenger, Gollwitzer, & Oettingen, 2013; Schouppe, De Houwer, Ridderinkhof, & Notebaert, 2012). For example, Aarts et al. (2012) demonstrated that false alarms in Go/noGo task lead to a faster evaluative categorization of negative words. More directly, Fritz and Dreisbach (2013) found that even in the absence of any action incongruent Stroop stimuli decrease subsequent evaluative ratings of neural stimuli, such as words or Chinese ideograms. In line with these findings Martiny-Huenger et al. (2013) utilized Eriksen flanker task (Eriksen & Eriksen, 1974) to show that conflict created by incongruent stimuli leads to less positive ratings for distractors. Finally, Chetverikov & Kristjánsson studied how the repetition benefits and switch costs in a color singleton visual search task are reflected in free choice biases and liking (Chetverikov & Kristjánsson, 2014). Their findings show that while distractor repetitions create bias in favor of target stimuli, liking ratings are decreased in the case of distractor-to-target switch. These studies show that conflict can be a source of negative affect.

It is quite possible that our results can be explained in terms of conflict. The question is then, what kind of conflict is it. Earlier, we suggested that one of the ways of testing the hypotheses in the absence of external feedback is to utilize parallel mechanisms that provide independent assessment of hypothesis correctness (Chetverikov, 2014). The idea is similar to Ramachandran (1990) suggestion that perception is a "bag of tricks" each of which is inefficient by itself, but together they can provide accurate estimates of the perceived. Inconsistency of hypotheses testing results provided by these parallel mechanisms can be considered as a conflict. On the other hand, it is also possible to imagine some other type of conflict, such as a conflict between competing accumulators collecting data in favor of each alternative. The notion of conflict may have different interpretations (Egner, 2008) and further studies may clarify this issue by comparing errors associated with different types of conflicts.

6.3. Affective feedback and aesthetic perception

The proposed account also has close parallels in the literature on aesthetic perception, both in visual art and in music literature (Huron, 2006; Leder, Belke, Oeberst, & Augustin, 2004; Ramachandran & Hirstein, 1999; Van de Cruys & Wagemans, 2011; Vuust & Kringelbach, 2010). Ramachandran and Hirstein (1999) proposed that we experience positive or negative aesthetic emotions because objects of art provide support for our perception of the world in an exaggerated form. Our perceptual system selects the most important features of objects. When the essence of an object is presented in art, it agrees with our internal model more than the real object and we like it more. For example, we like Nixon's caricature because it provides support for our perceptual model of Nixon even more than his photograph. In a similar fashion, Van de Cruys and Wagemans (2011) used a predictive coding framework to suggest that art appreciation is determined by the reduction of prediction error, often initially increased by artist's manipulations. For example, when one looks at Picasso's cubist painting the initial predictions about the depicted objects fail because the artist purposely separates the object parts, turns curves into sharp angles, distorts perspective, etc. However, if one manages to predict correctly the meaning of a painting and the reasons behind the artistic distortions, then the positive gain from finding support for this prediction leads to an overall positive evaluation.

6.4. Conclusions

The obtained findings are in accordance with the findings of Muth and Carbon (2013) by showing that perceptual insights are intrinsically reinforced with positive affect. This can explain why it is hard to "unsee" a Gestalt once you grasped it. All the other perceptual hypotheses will be inconsistent with the obtained interpretation and will evoke negative affect. Muth and Carbon (2013) discuss this finding in relation to art perception and art evaluation. However, we argue that the obtained findings have wider implications. On a more general level, we suggest that one of the means by which people discriminate between correct and incorrect perceptual hypotheses is by relying on affective feedback. Thus, we might be able to discern between a wife and a hat simply by analyzing the affective feedback that is evoked by each of these hypotheses.

Acknowledgments

This study has been supported in parts by the Russian Foundation for Humanities (#12-36-01294) and Saint-Petersburg State University (research grant 8.38.287.2014). The authors are also grateful to Árni Kristjánsson for the help in editing and insightful comments.

References

Aarts, K., De Houwer, J., & Pourtois, G. (2012). Evidence for the automatic evaluation of self-generated actions. *Cognition*, *124*(2), 117–127. http://dx.doi.org/10.1016/j.cognition.2012.05.009.

Aarts, K., De Houwer, J., & Pourtois, G. (2013). Erroneous and correct actions have a different affective valence: Evidence from ERPs. *Emotion (Washington, D.C.)*, *13*(5), 960–973. http://dx.doi.org/10.1037/a0032808.

Allakhverdov, V. M., & Gershkovich, V. A. (2010). Does consciousness exist? — In what sense? *Integrative Psychological & Behavioral Science*, *44*(4), 340–347. http://dx.doi.org/10.1007/s12124-010-9133-8.

Baayen, R. H., Davidson, D. J., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.0–4 (Retrieved from http://cran.r-project.org/package=lme4).

Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluence/attributional model of the mere exposure effect. *Social Cognition*, *12*(2), 103–128.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152. http://dx.doi.org/10.1037/h0043805.

Chetverikov, A. (2013). Automatic influence of irrelevant affect on confidence judgments. *Perception*, *42*, 242 (ECVP Abstract supplement).

Chetverikov, A. (2014). Warmth of familiarity and chill of error: Affective consequences of recognition decisions. *Cognition & Emotion*, *28*(3), 385–415. http://dx.doi.org/10.1080/02699931.2013.833085.

Chetverikov, A., Jóhannesson, Ó., & Kristjánsson, Á. (2014). *Blaming the victims of your own mistakes: How the accuracy of visual search influences evaluation of stimuli.* (submitted for publication).

Chetverikov, A., & Kristjánsson, Á. (2014). *Look out! Monsters!: History effects in visual search as reflected in choice biases and liking.* (submitted for publication).

Dan-Glauser, E. S., & Scherer, K. R. (2011). The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. *Behavior Research Methods*, *43*(2), 468–477. http://dx.doi.org/10.3758/s13428-011-0064-1.

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *Brain and Cognition*, *78*(2), 94–98. http://dx.doi.org/10.1016/j.bandc.2011.12.003.

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*(10), 374–380. http://dx.doi.org/10.1016/j.tics.2008.07.001.

Eriksen, B.A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. http://dx.doi.org/10.3758/BF03203267.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. *Research Methods in Personality and Social Psychology*, *11*, 74–97.

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nature reviews. Neuroscience*, *11*(2), 127–138. http://dx.doi.org/10.1038/nrn2787.

Fritz, J., & Dreisbach, G. (2013). Conflicts as aversive signals: Conflict priming increases negative judgments for neutral stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(2), 311–317.

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*(1358), 1121–1127. http://dx.doi.org/10.1098/rstb.1997.0095.

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*(2), 189–197. http://dx.doi.org/10.1016/j.bandc.2003.11.001.

Hohwy, J. (2012, April). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*. http://dx.doi.org/10.3389/fpsyg.2012.00096.

Hohwy, J. (2013). Functional integration and the mind. *Synthese*, *159*(3), 315–328. http://dx.doi.org/10.1007/sl.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. http://dx.doi.org/10.1037/a0028347.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80–113. http://dx.doi.org/10.1037/a0025648.

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, *95*(4), 489–508. http://dx.doi.org/10.1348/0007126042369811.

Luu, P., Collins, P., & Tucker, D.M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology. General*, *129*(1), 43–60.

Martiny-Huenger, T., Gollwitzer, P.M., & Oettingen, G. (2013). Distractor devaluation in a flanker task: Object-specific effects without distractor recognition memory. *Journal of Experimental Psychology. Human Perception and Performance*. http://dx.doi.org/10.1037/a0034130 (Advance online publication).

Muth, C., & Carbon, C. -C. (2013). The aesthetic aha: On the pleasure of having insights into Gestalt. *Acta Psychologica*, *144*(1), 25–30. http://dx.doi.org/10.1016/j.actpsy.2013.05.001.

Oppenheimer, D.M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. http://dx.doi.org/10.1016/j.tics.2008.02.014.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. http://dx.doi.org/10.1037/a0019737.

Ramachandran, V. S. (1990). Visual perception in people and machines. In R. Blake, & T. Troscianko (Eds.), *AI and the eye* (pp. 21–77). New York: Wiley.

Ramachandran, V. S., & Hirstein, W. (1999). The science of art. *Journal of Consciousness Studies*, *6*(6–7), 15–51.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697–719. http://dx.doi.org/10.1037/a0033152.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–382. http://dx.doi.org/10.1207/s15327957pspr0804_3.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45–48.

Sacks, O. (2011). *The man who mistook his wife for a hat.* Pan Macmillian.

Schouppe, N., De Houwer, J., Ridderinkhof, K. R., & Notebaert, W. (2012). Conflict: Run! Reduced Stroop interference with avoidance responses. *Quarterly Journal of Experimental Psychology (2006)*, *65*(6), 1052–1058. http://dx.doi.org/10.1080/17470218.2012.685080.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of Weil-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513–523.

Starcevic, V., & Berle, D. (2006). Cognitive specificity of anxiety disorders: A review of selected key constructs. *Depression and Anxiety*, *23*, 51–61. http://dx.doi.org/10.1002/da (August 2005).

Tolin, D. F., Abramowitz, J. S., Brigidi, B.D., Amir, N., Street, G. P., & Foa, E. B. (2001). Memory and memory confidence in obsessive–compulsive disorder. *Behaviour Research and Therapy*, *39*(8), 913–927.

Van de Cruys, S., & Wagemans, J. (2011). Putting reward in art: A tentative prediction error account of visual art. *I-Perception*, *2*(9), 1035–1062. http://dx.doi.org/10.1068/i0466aap.

Von Helmholtz, H. (1866). Concerning the perceptions in general. (3rd ed.). *Treatise on physiological optics*, *Vol. III*, (translated by J. P. C. Southall 1925 Opt. Soc. Am. Section 26, reprinted New York: Dover, 1962).

Vuust, P., & Kringelbach, M. L. (2010). The pleasure of making sense of music. *Interdisciplinary Science Reviews*, *35*(2), 166–182. http://dx.doi.org/10.1179/030801810X12723585301192.

Wiswede, D., Münte, T. F., Goschke, T., & Rüsseler, J. (2009). Modulation of the error-related negativity by induction of short-term negative affect. *Neuropsychologia*, *47*(1), 83–90. http://dx.doi.org/10.1016/j.neuropsychologia.2008.08.016.