Journal of Experimental Psychology: General

The Architecture of Prototype Preferences: Typicality, Fluency, and Valence

Tobias Vogel, Moritz Ingendahl, and Piotr Winkielman

Online First Publication, June 25, 2020. http://dx.doi.org/10.1037/xge0000798

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

http://dx.doi.org/10.1037/xge0000798

BRIEF REPORT

The Architecture of Prototype Preferences: Typicality, Fluency, and Valence

Tobias Vogel and Moritz Ingendahl
University of Mannheim

Piotr Winkielman
University of California, San Diego, and SWPS University of Social Sciences and Humanities

A classic phenomenon known as prototype preference effect (PPE) or beauty-in-averageness effect is that prototypical exemplars of a neutral category are preferred over atypical exemplars. This PPE has been explained in terms of deviance avoidance, hedonic fluency, or preference for certainty and familiarity. However, typicality also facilitates greater activation of category-related information. Thus, prototypes rather than atypical exemplars should be more associated with the valence of the category, either positive or negative. Hence, we hypothesize that the evaluation of a prototype depends on the valence of its category. Results from three experiments crossing a standard PPE paradigm with an evaluative conditioning procedure support our hypothesis. We show that for positive categories, greater typicality increases liking. Critically, for negative categories, greater typicality decreases liking. This pattern of results challenges dominant explanations of prototype evaluation.

Keywords: beauty-in-averageness, categorization, evaluative conditioning, fluency, valence

Supplemental materials: http://dx.doi.org/10.1037/xge0000798.supp

A classic phenomenon in psychology is the *prototype preference effect (PPE)*: preference for typical category exemplars over atypical exemplars. Closely related is the beauty-in-averageness effect: preference for the category average over its constituent exemplars (Galton, 1879; Langlois & Roggman, 1990). The PPE occurs with natural and artificial stimuli, across different populations, ages, and even species. As reviewed next, the phenomenon has been explained by multiple influential frameworks across psychology. Building on categorization theory, we present an account in which the PPE crucially depends on the valence of attributes associated with prototypes. This account generates a novel empirical prediction—a reversal of the classic effect, with *lower* preference for typical than atypical exemplars within negatively valenced categories.

Tobias Vogel and Moritz Ingendahl, Department of Consumer and Economic Psychology, University of Mannheim; Piotr Winkielman, Department of Psychology, University of California, San Diego, and Department of Psychology, SWPS University of Social Sciences and Humanities.

The authors want to thank Andy Arnold for helpful comments on the article. Additional information is available at https://osf.io/qyf58/. Findings from Experiment 1 have been presented at the Tagung experimentell arbeitender Psychologen (TeaP) 2017 in Dresden, Germany.

Correspondence concerning this article should be addressed to Tobias Vogel, Department of Consumer and Economic Psychology, University of Mannheim, A5, 6 B117, 68159 Mannheim, Germany. E-mail: tobias.vogel@uni-mannheim.de

Prototype Evaluation: Major Explanations

PPE From Deviance Detection

Evolutionary accounts suggest that the PPE results from "deviance detection"—negative sexual selection for atypical organisms (Langlois & Roggman, 1990; Rhodes & Tremewan, 1996; Symons, 1979; Thornhill & Gangestad, 1993). However, the PPE also occurs with objects (e.g., Halberstadt & Rhodes, 2003), including consumer goods (e.g., Landwehr, Labroo, & Herrmann, 2011), and even abstract graphical patterns (e.g., Winkielman, Halberstadt, Fazendeiro, & Catty, 2006). Accordingly, many frameworks focus on basic cognitive and affective processes.

PPE From Preference for Certainty, Familiarity, or Efficient Coding

Note that the aggregate similarity to all category exemplars is highest for prototypes and they are most unambiguously identified as category members (Markman & Ross, 2003). As such, prototypes are the most certain, most "familiar" category members. The PPE could then occur because uncertainty triggers negative affect (Friston, Adams, & Montague, 2012; Hsu & Preuschoff, 2015; Klein, Cosmides, Tooby, & Chance, 2002). Or because familiarity triggers positive affect (Titchener, 1915). Related arguments emphasize that prototypes are statistically typical and thus more efficiently coded (Dotsch, Hassin, & Todorov, 2016; Vogel, Carr, Davis, & Winkielman, 2018). Thus, prototypes could be preferred because they are simpler and less energy-demanding (Ryali & Yu, 2018).

PPE From Hedonic Fluency

This explanation derives from observations that prototypical stimuli are classified faster than atypical ones (e.g., Posner & Keele, 1968). Could classification ease be the source of prototype preference? This is suggested by research indicating that many ways of increasing processing ease (i.e., fluency) enhance stimulus evaluation, perhaps because fluency is a cue to successful pattern recognition, higher familiarity, or comes with lower processing costs (Reber, Schwarz, & Winkielman, 2004; Winkielman, Schwarz, Fazendeiro, & Reber, 2003). This "hedonic fluency" framework proposes that prototypes are attractive partly because they are fluent. In one supporting experiment, participants first studied several exemplars (dot patterns) from two categories, "Acks" and "Blubs" (Winkielman et al., 2006). Participants were then tested on patterns that contained exemplars, but also the never-presented prototype (i.e., the category average). Participants classified and evaluated the test stimuli. The results showed that typicality positively predicted classification fluency, which in turn positively predicted stimulus evaluation (see also Halberstadt & Winkielman, 2014). Thus, the PPE occurs even with highly controlled, unfamiliar, and arbitrary stimuli, and is at least partly due to prototype fluency.

Informational Function of Prototypes

The just discussed perspectives suggest that prototypes should always be preferred over atypical exemplars. However, they neglect a key function of categories—predicting the presence of related attributes. Knowing that raspberries belong to the category of fruit, one can infer that they are sweet, which may then drive approach behaviors (e.g., consumption). Importantly, category properties are more likely to be assigned to more typical exemplars (Rosch, 1975; Rosch & Mervis, 1975). For example, property "sweet" is more likely to be assigned to a more typical fruit. Of course, a category can be associated with negative attributes (e.g., hooligans with aggressiveness), and a more typical exemplar will more likely be assigned more negative attributes. Thus, one should predict a reversal of PPE with negative categories, because prototypes, as opposed to atypical exemplars, are more likely to activate attributes stored at the category level (cf. Nedungadi, 1990). In other words, prototypes "inherit" more goodness of the positive category and more badness of the negative category than atypical exemplars. Note that in this model, fluency still is a valid cue to category membership, because typical items are more fluent (Posner & Keele, 1968). In fact, enhancing fluency increases classification of items as category members (Oppenheimer & Frank, 2008; Whittlesea & Leboe, 2000). Critically, in this model fluency can signal typicality for positive as well as negative categories.

Present Research and Hypotheses

Here we test the influence of category valence on prototype evaluation. The dominant accounts (deviance detection, certainty preference, hedonic fluency) predict that prototypicality will always enhance stimulus evaluation. Thus, these accounts predict only a main effect of typicality. In contrast, the account emphasizing the informational functions of categories suggests that the category valence will moderate the PPE. We therefore predict an interaction of category valence and stimulus typicality on evaluations. Furthermore, this account suggests that participants will use classification fluency as a cue to category typicality, but that for negative categories higher fluency will be associated with lower evaluation.

Note, however, that we do not necessarily rule out some positivity deriving from typicality, certainty, or hedonic marking of fluency. The key point is that those alternative accounts do not specifically predict a Typicality × Valence interaction. However, their discussed mechanisms might still contribute to overall stimulus liking. Thus, it is an open question if the interaction will occur in addition to a typicality main effect (thus, a hybrid interaction) or if it will attenuate the classic typicality main effect (indicating a cross-over interaction, including a reversal of the PPE for negative categories). We return to the complex interplay of these mechanisms and their statistical implications in the discussion.

To test our hypothesis, we combine an established paradigm from PPE research (Winkielman et al., 2006) and an established procedure from evaluative conditioning (EC) research (Baeyens, Eelen, & Van den Bergh, 1990; Staats & Staats, 1958). We discuss the theoretical interpretation of the EC procedure later, but essentially it aims to create a change in liking of a conditioned stimulus (CS) due to its paring with a valenced unconditioned stimulus (US; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Thus, in a learning phase, exemplars of two different categories (CS) are paired with valenced stimuli (US), either positive or negative. In a test phase, participants indicate their liking for some exemplars as well as the prototype stimulus, which was not shown during the learning phase.

In the first two studies, we show that the PPE is qualified by category valence for conditions of supervised learning (Experiment 1) and incidental learning (Experiment 2). Experiment 3 tests the mediating role of classification fluency.

Experiment 1

As in Winkielman et al. (2006), participants were shown abstract dot patterns varying in typicality of their category, and were also simultaneously exposed to images with positive or negative content.

Method

Design and participants. Ninety-seven students from the University of Mannheim (77 female; 20 male; $M_{\text{age}} = 22.26$, $SD = 4.82$) participated for course credit and biscuits. The sample size based on an a priori power analysis for a repeated-measures analysis of variance with $f = .2$; $df_{\text{num}} = 2$; $\alpha = .05$; $1 - \beta = .8$, which yielded a required sample size of 84 participants. The design was within-subjects: a 2 (US Valence: positive vs. negative) × 3 (Test Stimulus: prototype vs. shown exemplar vs. unshown exemplar). One participant was excluded (he mixed up the scale anchors).

Materials and procedure. The study was run on computers in groups up to four participants. In total, participants completed six blocks, each starting with a learning phase followed by a test phase. In the learning phase, participants saw a succession of 56 trials, in which 28 CS from each of two categories called "Acks" and "Blubs" were presented simultaneously with valenced stimuli (cf. Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012). Exemplars from one category were consistently paired with positive, whereas exemplars from the other were consistently paired with negative stimuli. In the test phase, participants were then asked to indicate their liking for three stimuli per category: the prototype (not shown in the learning phase), and two exemplars - one shown and the other not shown in the learning phase. Specifically, for each stimulus, they indicated an evaluation using the number 1 (*do not like at all*) to 9 (*like a lot*) on the keyboard. After the evaluation phase, participants then proceeded with the next block. Upon completion of the six blocks, participants indicated demographic data, and were debriefed.

CS stimuli. Dot patterns served as stimuli and were generated before each block following the algorithm from Winkielman et al. (2006). First, a prototype pattern for each category was generated by randomly assigning eight dots to a 30 × 30 grid. Next, distortions of this pattern were then created to obtain exemplars of each category (Figure 1a). To do so, each dot of the prototype pattern was moved with a certain probability by either 1-, 2-, 3-, or 4-dot diameters. We realized four levels of distortion, with higher levels reflecting higher probabilities that a dot of the original pattern would be moved (Figure 1; for the probabilities, see Table 1 in Winkielman et al., 2006). For each category, we obtained seven dot pattern distortions per level, yielding 28 individual exemplars per category. In the learning phase, exemplars from two categories were shown, thus a total of 56 exemplars each presented once together with a co-occurring US.

US stimuli. We used images from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). That is, on each trial of the learning phase, a picture was randomly selected from a pool of 42 pictures with positive valence, $M = 7.37$, $SD = .90$, or negative valence, $M = 2.64$, $SD = .38$, depending on category valence condition. Thus, across 56 trials, one category was paired with multiple positive IAPS pictures, but the other category was paired with multiple negative IAPS pictures (Figure 1b).

Results

Analyses used a 2 (US Valence: positive vs. negative) × 3 (Test Stimulus: prototype vs. shown exemplar vs. unshown exemplar) analysis of variance, run in the GENLINMIXED procedure in SPSS Version 25.0. Due to the nested data structure, we used a multilevel mixed-modeling approach (Snijders & Bosker, 2012), with participants treated as Level 2, stimuli treated as Level 1 units. The maximal model reaching convergence (Barr, Levy, Scheepers, & Tily, 2013) was a random-intercept model with trial order as repeated measure (see S2 in the online supplemental materials for details).

Analyses yielded a main effect of US valence, $F(1, 3450) = 96.40$, $p < .001$. Stimuli from the category paired with positive US were rated more favorably than stimuli from the category paired with negative US. Interestingly, there was no main effect for test stimulus type, $F(2, 3450) = 0.68$, $p = .506$, offering no evidence for a general PPE. Instead, the critical interaction was significant, $F(2, 3450) = 7.19$, $p = .001$. Figure 2 shows that the PPE occurred only for categories paired with positive US. Indeed, pairwise contrasts indicate that prototypes, $M = 6.30$, $SE = 0.14$, were preferred over unshown exemplars, $M = 6.00$, $SE = 0.13$; $t(3450) = 3.40$, $p = .001$, $d = 0.13$, as well as over shown exemplars, $M = 6.07$, $SE = 0.13$; $t(3450) = 2.63$, $p = .008$, $d = 0.10$. However, negative US prototypes, $M = 4.01$, $SE = 0.15$, were actually liked less than unshown exemplars, $M = 4.20$, $SE = 0.13$; $t(3450) = -2.29$, $p = .022$, $d = -0.08$, or shown exemplars, $M = 4.13$, $SE = 0.12$; $t(3450) = -1.42$, $p = .156$, $d = -0.05$, though only the former contrast was significant.

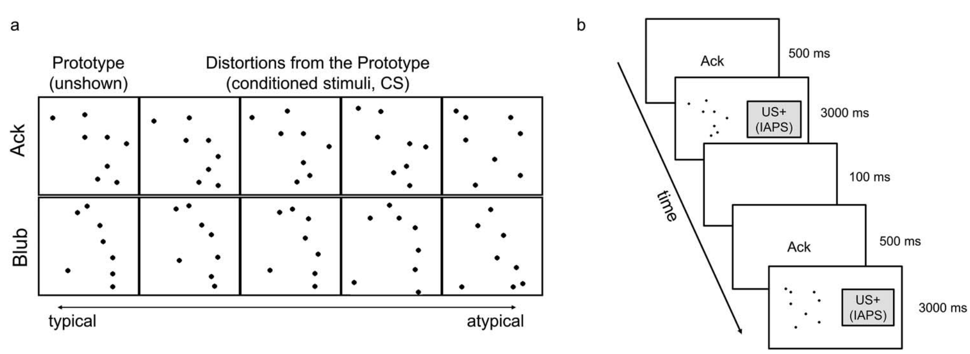

Figure 1. Sample (a) stimuli and (b) procedure used in the experiments. Sample stimuli of category exemplars used in the learning phase are shown in (a). They were obtained by distorting the prototype (left column). The level of distortion (1–4) increases from the second to the last column, thus decreasing typicality. The learning phase is illustrated in (b). For supervised learning (Experiments 1 and 3), a trial started with the respective category label, "Ack" or "Blub" (500 ms). This was followed by the presentation of a CS and an unconditioned stimulus (US; 3,000 ms), which either appeared on the right or left side of the dot pattern. Then there was an intertrial blank screen (100 ms). For unsupervised learning (Experiment 2), the trial began with a fixation cross instead of a category label. Patterns of one category were paired with different positive images (US+, e.g., a puppy), but patterns of the other category were paired with different negative images (US–, e.g., a dental treatment). The prototype was not conditioned during the learning phase, but served as test stimulus in the evaluation phase.

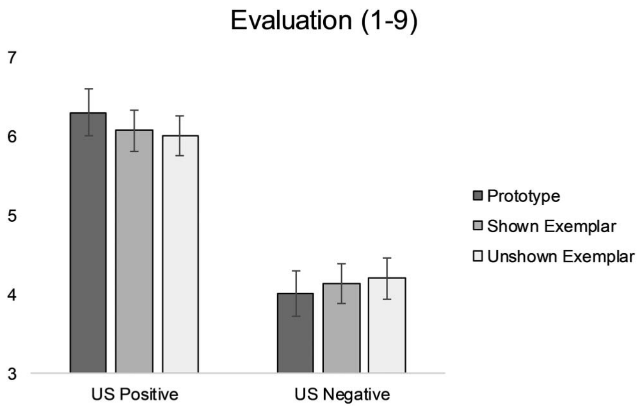

Figure 2. Results from Experiment 1. Error bars indicate 95% confidence intervals. US = unconditioned stimulus.

Discussion

Effects of typicality on liking were moderated by US valence. The PPE occurred only for categories associated with positive US. For categories associated with negative US, prototypes received lower evaluations as compared to less typical exemplars. The present results fit the idea that prototypes activate category-related content.

Note that participants were explicitly informed about the belongingness of each exemplar to a category. Yet, the PPE also occurs for unsupervised incidental learning, without any reference to a category (e.g., Vogel et al., 2018; Experiment 3). However, it is possible that for incidental learning, prototypes do not activate category valence.

Experiment 2

Experiment 2 tested whether category valence moderates the PPE even for unsupervised category learning.

Method

Design and participants. Eighty students from the University of Mannheim (70 female; 10 male; $M_{\text{age}} = 20.31$, $SD = 3.17$) participated for course credit or money (€4, approximately US$4.50). The design was within-subject: a 2 (US Valence: positive vs. negative) × 3 (Test Stimulus: prototype vs. shown exemplar vs. unshown exemplar).

Materials and procedure. Materials in Experiment 2 were identical to Experiment 1, except for the following. Instructions did not refer to categories or categorization, nor were there any category labels displayed during the learning phase. Also, the item pool for the USs was increased to 50 images per valence condition.

Results

We used the same analysis as in Experiment 1. Again, it yielded a significant effect of US valence, $F(1, 2874) = 157.11$, $p < .001$, but no effect for the test stimulus, $F(2, 2874) = 0.14$, $p = .862$. The interaction was significant, again, $F(2, 2874) = 10.74$, $p < .001$. As shown in Figure 3, prototypes, $M = 6.22$, $SE = 0.15$, were preferred over unshown exemplars, $M = 5.81$, $SE = 0.14$; $t(2874) = 3.60$, $p < .001$, $d = 0.16$, and over shown exemplars, $M = 5.96$, $SE = 0.13$; $t(2874) = 2.38$, $p = .018$, $d = 0.10$, when categories were paired with positive US. Again, pairings with a negative US yielded a reversal: Prototypes were evaluated lower, $M = 3.35$, $SE = 0.15$, than unshown exemplars, $M = 3.68$, $SE = 0.13$; $t(2874) = -3.27$, $p = .001$, $d = -0.13$, or shown exemplars, $M = 3.60$, $SE = 0.13$; $t(2874) = -2.23$, $p = .026$, $d = -0.10$.

Discussion

Again, the valence associated with the category moderated the PPE, with PPE reversed for categories paired with negative valence. Yet, it is still an open question whether category fluency plays a role in these effects. Experiment 3 was designed to address this.

Experiment 3

Experiment 3 measured fluency before the test phase using a classification task. We predict that typicality increases fluency, which should enhance evaluation for positive categories but lower it for negative categories.

Method

Design and participants. Fifty-seven participants (15 male; 42 female, $M_{\text{age}} = 23.49$, $SD = 7.32$) from the Universities of Heidelberg and Mannheim were paid €4. The design was within-subjects: a 2 (Valence) × 5 (Typicality).

Materials and procedure. The procedure was the same as in Experiment 1, except for the following. First, after the learning phase, and before the evaluation phase, we inserted a classification phase. In this phase, participants saw the test stimuli in random order. For each stimulus, they had to decide as fast as possible whether it belonged to the category of “Acks” or “Blubs,” by pressing “A” or “K,” on the keyboard, where labels were randomly assigned. The resulting reaction times (RTs) were trimmed (200 ms $< \text{RT} <$ 3,000 ms; see Vogel et al., 2018). RTs served as an indicator of fluency, with lower scores reflecting higher fluency. Second, to allow for a reliable RT measurement, we increased the number of test stimuli. We asked about five stimuli per category (prototype, as well as exemplars of each distortion level, 1, 2, 3, and 4, all of which were not shown in the learning phase). In subsequent

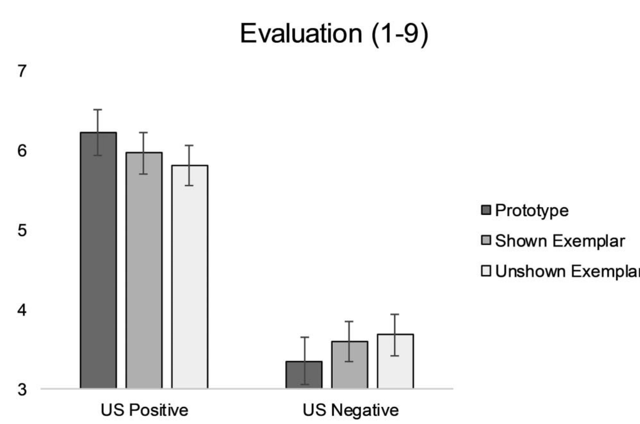

Figure 3. Results from Experiment 2. Error bars indicate 95% confidence intervals. US = unconditioned stimulus.

analyses, typicality is therefore treated as continuous predictor, ranging from 0 (*highest level of distortion*) to 4 (*prototype*).

Results

Evaluation scores were subjected to a multilevel mixed model regression and were predicted from US valence (positive = 1; negative = −1), typicality, and their interaction. Only random intercepts were modeled using the R package lme4 (Version 1.1–21; Bates, Mächler, Bolker, & Walker, 2015) as models including slopes for valence or typicality failed to converge. Besides a significant intercept, $b = 4.83$, $SE = 0.14$; $t(72.84) = 34.80$, $p < .001$, there was a significant effect of US valence, $b = 0.89$, $SE = .06$; $t(3230.40) = 14.92$, $p < .001$. While the linear effect of typicality was not significant, $b = 0.02$, $SE = .02$; $t(3230.68) = 0.69$, $p = .488$, a significant effect emerged for the critical Typicality × Valence interaction, $b = 0.08$, $SE = 0.02$; $t(3230.50) = 3.18$, $p = .001$ (Figure 4a). Simple slope analyses show that typicality increases liking for positive US, $b = 0.09$, $SE = 0.03$; $t(3230.67) = 2.75$, $p = .006$, but decreases liking for negative USs, $b = -0.06$, $SE = 0.03$; $t(3230.51) = -1.76$, $p = .079$, though the latter trend was not significant.

Next, classification times were analyzed analogously. The intercept was at, $b = 955.79$ ms, $SE = 30.56$ ms; $t(78.21) = 32.59$, $p < .001$. Neither valence, $b = 5.42$ ms, $SE = 14.81$ ms; $t(3230.04) = 0.37$, $p = .715$, nor the interaction were significant, $b = -3.36$ ms, $SE = 6.04$ ms; $t(3230.171) = -0.56$, $p = .578$, but the effect of typicality was: Response times decrease with higher typicality, $b = -32.12$ ms, $SE = 6.04$ ms; $t(3230.41) = -5.32$, $p < .001$ (Figure 4b).

Finally, we tested if fluency accounts for the effects of typicality and US valence on evaluations. A multilevel moderated mediation analysis was carried out using the R-package "lavaan" (Version 0.6–5; Rosseel, 2012). We tested whether typicality reduces classification times, which in turn should enhance evaluations for positively, but lower evaluations for negatively conditioned categories. Supporting our expectations, the index of moderated mediation was significant, $b = 0.02$, 95% confidence interval (CI) [0.006, 0.024], $p = .001$.

For positive categories, the indirect effect of typicality via fluency on liking was positive and significant, $b = 0.02$, 95% CI [0.004, 0.026], $p = .003$. Yet, fluency did not fully account for the effect as is evident from a significant direct effect of typicality on liking, $b = 0.08$, 95% CI [0.021, 0.133], $p = .007$. Crucially, the indirect effect of typicality via fluency on liking reversed for negative categories, $b = -0.01$, 95% CI [−0.026, −0.003], $p = .010$. Thus, same as for positive categories, typicality increased fluency. However, fluency in turn decreased rather than increased liking. The direct effect was not significant for negative categories, $b = -0.05$, 95% CI [−0.103, 0.008], $p = .094$.

Discussion

Experiment 3 confirmed that the influence of typicality on liking depends on category valence. It also showed that fluency mediates the effects. Crucially, the effect of fluency on liking is not always positive—for negative categories, higher fluency goes with lower liking, presumably reflecting ease of activation for properties of negative category.

General Discussion

Three experiments show that category valence moderates the effect of typicality on liking. For positive categories, we found the classic preference for typical over atypical exemplars. For negative categories the opposite pattern emerged—typical members were liked less than atypical members. This novel effect occurred robustly across learning conditions—supervised and unsupervised category learning. As such, the results support the account emphasizing the informational function of prototypes. This model is based on research showing that perceivers extract category prototypes in similar paradigms (Smith & Minda, 2002) and that category attributes are more likely to be assigned to typical versus atypical category members (Rosch, 1975; Rosch & Mervis, 1975). Consistently, Experiment 3 showed that classification fluency contributes to the effect. Mediation analyses revealed that for positive categories, typicality increased fluency. This, in turn, yielded higher stimulus evaluation. Typicality also increased fluency for negative categories. However, in this case higher fluency decreased stimulus evaluation. This suggests that prototypes facilitate the activation of category knowledge in memory and inherit more of category valence. Note also that this is true even though

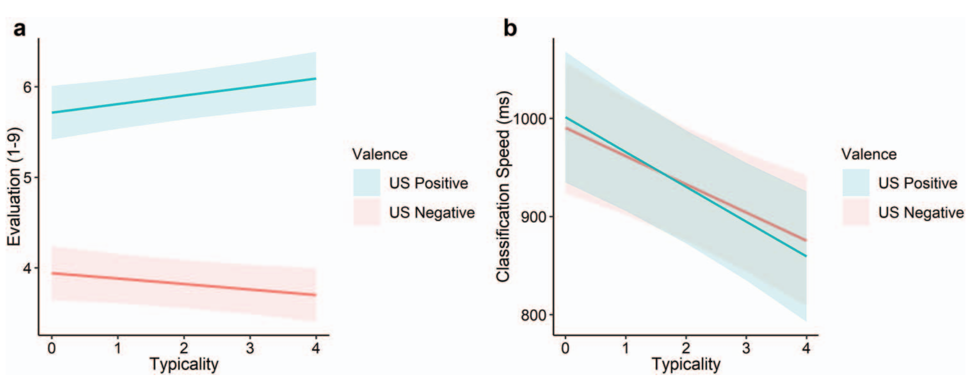

Figure 4. Results from Experiment 3: (a) Evaluations and (b) classification times predicted from typicality, valence, and their interaction. Typicality ranged from 0 (*highest level of distortion*) to 4 (*prototype*). Shaded areas represent 95% confidence intervals. US = unconditioned stimulus. See the online article for the color version of this figure.

prototypes were never encountered in the study phase, let alone been paired with valenced stimuli.

The present pattern of findings cannot be explained by models proposing a general dislike for "deviants" (Symons, 1979), or a general preference for certainty (Hsu & Preuschoff, 2015), familiarity (Titchener, 1915) or statistical typicality (Ryali & Yu, 2018). They also challenge any strong claims from the hypothesis about hedonic marking of fluency (Winkielman et al., 2003). After all, typicality robustly increased fluency, but failed to produce a positive main effect on evaluations, regardless of category valence (but see next). But why is there such a rich literature showing the classic PPE for neutral categories? There are multiple possibilities.

First, one can think of the standard PPE as a generalized EC effect. Note that the default valence in our environments is (somewhat) positive and positive experiences are more frequent that negative ones (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008). Accordingly, on repeated encounters, a class of new neutral stimuli becomes associated with this contextual positive valence. In fact, this "safe-context learning" is one explanation of the mere-exposure effect (Zajonc, 2001). As such, the PPE with neutral stimuli could be explained with a learning approach, and the hedonic marking of fluency may reflect the ease with which positive memory content can be activated. Ceteris paribus, a stimulus will activate positively valenced content in memory, unless distinct negative associations exist (Alves et al., 2015). However, without additional assumptions (e.g., stronger EC effects for negative than positive USs), the generalized EC effect would also predict a typicality main effect which we did not observe.

Similar considerations pertain to a model assuming independent processes of positive fluency and conditioned valence. In fact, this idea of two independent processes has been proposed by Landwehr, Golla, and Reber (2017). Their research showed that negative USs led to a slight decrease in evaluation of repeatedly paired CSs. However, not only positive, but also neutral USs led to an increase in CS evaluation. Thus, in total, they found a positive fluency effect (induced by CS repetition) in addition to an EC effect. Again, our data does not support the notion of a strong, separate additive positive contribution of fluency because we did not find any statistical evidence for a main effect. Moreover, we found that for negative categories higher fluency was associated with decreased liking. We return to the idea of a two-process model shortly.

As an alternative to the aforementioned models, one may speculate that positive valence from fluency can add to, but can also be attenuated by category valence. In line with this notion, Albrecht and Carbon (2014) reported that priming-induced fluency added positivity to the initial valence of a US if the fluency manipulation was strong (i.e., long prime durations, Experiment 2), but amplified evaluation of an initially positive or negative US if the fluency manipulation was weak (i.e., very short prime durations, Experiment 1). Similarly, in our case, a typicality main effect may be observed if the typicality manipulation is stronger. This, for instance, occurs when exemplars within categories are more distinct, but between-category variance decreases (Vogel et al., 2018). Likewise, a PPE main effect might become more likely if, at the judgment stage, the information about the earlier category valence manipulation is weaker, less certain, or less salient. In fact, EC effects seem to depend on the awareness of the valence the CS had been paired with (Hofmann et al., 2010). Accordingly, a prototype preference main effect is likely to be observed if people have unreliable memory for the negative US valence, or if they do not generate the valence association at all. This logic could solve the puzzle of why in Galton's (1879) a composite picture of criminals was liked more than their contributing criminal exemplars. This would occur if the composite picture looked sufficiently different from individual criminals, and similar to composite pictures of noncriminals (essentially obscuring the associated negative valence).

Moreover, one could speculate on temporal dynamics which integrates the informative value of fluency in a two-step process. A typical stimulus, which is easy to identify, initially generates some positive affect, as proposed in the hedonic fluency hypothesis. However, the subsequent identification of the stimulus as a category member necessarily yields the activation of respective category knowledge, which, as a stronger and more diagnostic cue, then overwrites any initial hedonic marking. According to that model, in valenced (but not neutral) categories, the affective reaction to the prototype is controlled by the more informative cue of category valence. This idea relates to the proposal that when more diagnostic valence cues are available, fluency is ignored as a hedonic cue (Schwarz, 2004). Relatedly, a process could start with an initial positive fluency experience but be followed by reappraisal of the fluency experience. Indeed, researchers have argued that fluency effects depend on naïve theories guiding context-dependent interpretations of the fluency experience (Alter & Oppenheimer, 2009; Schwarz, 2004). Accordingly, people engage in higher order cognitive processes and use the typicality and fluency experience depending on task affordances (Vogel, Silva, Thomas, & Wänke, 2020).[1] Future research may provide stringent tests for these questions, for example by manipulating the content of naïve theories (e.g., Briñol, Petty, & Tormala, 2006).

Lastly, future research may elaborate on the underlying categorization principles. For instance, categorization can be based on visual similarity (e.g., among exemplars or toward a prototype) or rules (e.g., defining features). Both can occur within the same individual (e.g., Smith, Patalano, & Jonides, 1998) and depend on stimulus complexity (Karlsson, Juslin, & Olsson, 2008). Whereas previous research suggests that difficult-to-verbalize dot patterns used in the present article are categorized based on visual similarity (cf. Vogel et al., 2018),[2] unidimensional stimuli may be categorized according to a simple rule (e.g., Category A members are bigger than Category B members). If so, category knowledge may

[1] To test for an adaptation to task affordances, we reanalyzed data from the present experiments by including the number of block to our main analyses (see S3 in the online supplemental materials for details). Indeed, in analyses in which the number of the block was entered as a continuous predictor, the three-way interaction turned out to be marginally significant in Experiment 1, and significant in Experiment 2, reflecting that the Typicality × Valence interaction effect tends to increase in the course of the experiment. However, the three-way interaction was not significant in Experiment 3. As such, our data suggests a training effect, which provides indirect evidence for a higher order cognitive process involved in category learning. Perhaps, participants interpret meta-cognitive experiences as a function of task requirements. However, this interpretation has to be met with caution, as it is post-hoc and only weakly supported statistically.

[2] Though the prototype has not been conditioned, each of its features (single dots) have been paired with valence multiple times. Thus PPE (and its reversals) could also reflect EC generalization at the feature level. One or the other, the present article indicates that the prototype of the negative category which has not been met before, is the one most likely to be avoided as it combines the features associated with negativity which are not necessarily present in an individual exemplar.

be most activated not by an average, but an extreme member (Davis & Love, 2010). For negative categories, we would therefore expect that the member that loads highest on the defining category dimension (e.g., the biggest Group A member) is evaluated most negatively, facilitating its avoidance (Kim & Murphy, 2011).

Conclusion

In conclusion, our research shows that for negative categories, the PPE can reverse, with typical exemplars liked less than atypical ones. The mechanisms and the boundary conditions of this reversal require further study, but it is an important revision of one of psychology classic effects.

References

Albrecht, S., & Carbon, C.-C. (2014). The fluency amplification model: Fluent stimuli show more intense but not evidently more positive evaluations. *Acta Psychologica, 148,* 195–203. http://dx.doi.org/10.1016/j.actpsy.2014.02.002

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13,* 219–235. http://dx.doi.org/10.1177/1088868309341564

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

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition and Emotion, 4,* 3–18. http://dx.doi.org/10.1080/02699939008406760

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Briñol, P., Petty, R. E., & Tormala, Z. L. (2006). The malleable meaning of subjective ease. *Psychological Science, 17,* 200–206. http://dx.doi.org/10.1111/j.1467-9280.2006.01686.x

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science, 21,* 234–242. http://dx.doi.org/10.1177/0956797609357712

Dotsch, R., Hassin, R. R., & Todorov, A. (2016). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1,* 1–6. http://dx.doi.org/10.1038/s41562-016-0001

Friston, K., Adams, R., & Montague, R. (2012). What is value-accumulated reward or evidence? *Frontiers in Neurorobotics, 6,* 11. http://dx.doi.org/10.3389/fnbot.2012.00011

Galton, F. (1879). Composite portraits, made by combining those of many different persons into a single resultant figure. *Journal of the Anthropological Institute of Great Britain and Ireland, 8,* 132–144. http://dx.doi.org/10.2307/2841021

Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin & Review, 10,* 149–156. http://dx.doi.org/10.3758/BF03196479

Halberstadt, J., & Winkielman, P. (2014). Easy on the eyes, or hard to categorize: Classification difficulty decreases the appeal of facial blends. *Journal of Experimental Social Psychology, 50,* 175–183. http://dx.doi.org/10.1016/j.jesp.2013.08.004

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136,* 390–421. http://dx.doi.org/10.1037/a0018916

Hsu, M., & Preuschoff, K. (2015). Uncertainty. In A. W. Toga (Ed.), *Brain mapping: An encyclopedic reference* (pp. 391–399). Cambridge, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-397025-1.00260-8

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141,* 539–557. http://dx.doi.org/10.1037/a0026477

Karlsson, L., Juslin, P., & Olsson, H. (2008). Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts? *Judgment and Decision Making, 3,* 244–260.

Kim, S., & Murphy, G. L. (2011). Ideals and category typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1092–1112. http://dx.doi.org/10.1037/a0023916

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2002). Decisions and the evolution of memory: Multiple systems, multiple functions. *Psychological Review, 109,* 306–329. http://dx.doi.org/10.1037/0033-295X.109.2.306

Landwehr, J. R., Golla, B., & Reber, R. (2017). Processing fluency: An inevitable side effect of evaluative conditioning. *Journal of Experimental Social Psychology, 70,* 124–128. http://dx.doi.org/10.1016/j.jesp.2017.01.004

Landwehr, J. R., Labroo, A. A., & Herrmann, A. (2011). Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science, 30,* 416–429. http://dx.doi.org/10.1287/mksc.1110.0633

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual* (Technical Report A-8). Gainesville: University of Florida.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science, 1,* 115–121. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00079.x

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129,* 592–613. http://dx.doi.org/10.1037/0033-2909.129.4.592

Nedungadi, P. (1990). Recall and consumer consideration sets: Influencing choice without altering brand evaluations. *The Journal of Consumer Research, 17,* 263–276. http://dx.doi.org/10.1086/208556

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition, 106,* 1178–1194. http://dx.doi.org/10.1016/j.cognition.2007.05.010

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353–363. http://dx.doi.org/10.1037/h0025953

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,* 364–382. http://dx.doi.org/10.1207/s15327957pspr0804_3

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science, 7,* 105–110. http://dx.doi.org/10.1111/j.1467-9280.1996.tb00338.x

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192–233. http://dx.doi.org/10.1037/0096-3445.104.3.192

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605. http://dx.doi.org/10.1016/0010-0285(75)90024-9

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Ryali, C., & Yu, A. J. (2018). Beauty-in-averageness and its contextual modulations: A Bayesian statistical account. *Advances in Neural Information Processing Systems*. Advance online publication. http://dx.doi.org/10.1101/360651

Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology, 14,* 332–348. http://dx.doi.org/10.1207/s15327663jcp1404_2

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65,* 167–196. http://dx.doi.org/10.1016/S0010-0277(97)00043-7

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 800–811. http://dx.doi.org/10.1037/0278-7393.28.4.800

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, United Kingdom: Sage.

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology, 57,* 37–40. http://dx.doi.org/10.1037/h0042782

Symons, D. (1979). *The evolution of human sexuality*. New York, NY: Oxford University Press.

Thornhill, R., & Gangestad, S. W. (1993). Human facial beauty: Averageness, symmetry, and parasite resistance. *Human Nature, 4,* 237–269. http://dx.doi.org/10.1007/BF02692201

Titchener, E. B. (1915). *A beginner's psychology*. New York, NY: Macmillan.

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: The density hypothesis. *Journal of Personality and Social Psychology, 95,* 36–49. http://dx.doi.org/10.1037/0022-3514.95.1.36

Vogel, T., Carr, E. W., Davis, T., & Winkielman, P. (2018). Category structure determines the relative attractiveness of global versus local averages. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 250–267. http://dx.doi.org/10.1037/xlm0000446

Vogel, T., Silva, R. R., Thomas, A., & Wänke, M. (2020). Truth is in the mind, but beauty is in the eye: Fluency effects are moderated by a match between fluency source and judgment dimension. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0000731

Whittlesea, B. W. A., & Leboe, J. P. (2000). The heuristic basis of remembering and classification: Fluency, generation, and resemblance. *Journal of Experimental Psychology: General, 129,* 84–106. http://dx.doi.org/10.1037/0096-3445.129.1.84

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes are attractive because they are easy on the mind. *Psychological Science, 17,* 799–806. http://dx.doi.org/10.1111/j.1467-9280.2006.01785.x

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Erlbaum.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10,* 224–228. http://dx.doi.org/10.1111/1467-8721.00154

Received October 5, 2019
Revision received March 31, 2020
Accepted April 12, 2020 ■