BRILL

Art & Perception 6 (2018) 41–66

brill.com/artp

Evaluation of Franz Marc's Color Theory using Implicit Testing Procedures

Annika Grotjohann and Daniel Oberfeld*

Institute of Psychology, Section Experimental Psychology, Johannes Gutenberg-Universität Mainz, Wallstr. 3, 55122 Mainz, Germany

Received 28 March 2017; accepted 27 December 2017

Abstract

In the early 20th century, the German expressionist painter Franz Marc formulated assumptions concerning the meanings of color, based on his individual sensations. He characterized the 'cool' blue as the 'masculine principle'. Yellow represented the 'feminine principle' which he declared as 'gentle, cheerful, and sensual'. This leaves red, the color he perceived as 'brutal and heavy'. Here, we tested some of the color–meaning associations assumed by Franz Marc via implicit measures based on response times, using Single Category Implicit Association Tests. The participants had to classify words as belonging to one of two semantic categories (e.g., masculine or feminine) by pressing one of two response buttons. One of the semantic categories shared a response button with a hue (e.g., masculine–blue), and this button needed to be pressed whenever a color patch was presented on the screen. The results showed that response times were faster when related hues and meaning categories (according to Marc's assumptions) shared the same response button, compared to when unrelated hues and meaning categories were assigned to the same button. The pattern of response times was compatible with the associations of blue–masculine, yellow–feminine, blue–cool and yellow–gentle as proposed by Marc. In addition, our data indicate associations of yellow–warm and red–warm, which were not explicitly formulated by Franz Marc. However, the proposed red–brutal association was not confirmed.

Keywords

Franz Marc, art, perception, color, emotion, color-meaning associations, hue, response time, Single Category Implicit Association Test

---

*To whom correspondence should be addressed. E-mail: oberfeld@uni-mainz.de

DOI: 10.1163/22134913-00002101

1. Introduction

1.1. Franz Marc's Color Theory

"Color, the fruit of light, is the foundation of painterly means of painting—and its language," as the French painter Robert Delaunay (1885–1941), a friend of Franz Marc, once said, emphasizing the great importance of colors and their function as a form of communication between the artist and the audience (Vriesen and Imdahl, 1967, p. 42; English translation by the authors). But how does the artist apply color? And does the audience understand intuitively what the artist intended to communicate?

Van Paasschen *et al.* (2014) provided evidence that artists are able to capture emotions in a set of low-level visual characteristics and that observers interpret them consistently (Melcher and Bacci, 2013). Evidently, painters know about such rules but they, and especially the expressionists like Marc and Delaunay, need a deeper knowledge about formal instruments such as colors and forms because their compositions were intended to communicate a message via pronounced contours and strong contrasts reflecting their subjective perception of reality (Fauth, 2010).

Franz Marc (1880–1916), a German expressionist painter who co-founded the famous art group 'Der Blaue Reiter' with his colleague Wassily Kandinsky, was very interested in theoretical aspects of art. He studied several color theories just to find that none of these worked perfectly for him and, as he put it, "My few superstitious notions about colors serve me better than all these theories" (Marc and Macke, 1964, p. 45; all translations from German to English by the authors). As Marc mainly painted animals and intended to make their soul visible (Partsch, 2005), he needed colors to represent their characteristics and emotions (see Partsch, 2010). In a letter to his friend, the painter August Macke, dating from 12 December 1910, he briefly formulated his own 'theory' concerning meanings of different primary hues:

> "Blue is the masculine principle, stern, and spiritual.
> Yellow the feminine principle, gentle, cheerful, and sensual.
> Red is matter, brutal and heavy and always the color which must be fought and vanquished by the other two!
> If, for instance, you mix the serious, spiritual blue with red, then you augment the blue to an unbearable mourning, and the reconciling yellow, the complementary to violet, will be indispensable (...). If you mix red and yellow to orange, you give the passive and feminine yellow a sensual power, for which the cool, spiritual blue will again be indispensable (...). But if you now mix blue and yellow to green, you bring red, the material, the earth to life (...)" (Marc and Macke, 1964, p. 28—see Note 1).

Thus, Franz Marc assumed blue to be associated with 'cool', although he did not state explicitly which color(s) he believed to represent 'warm'. In the following, we assumed that the two other primary colors mentioned by Marc, red and yellow, should be more strongly associated with ‘warm’ than with ‘cool’. An association between color and temperature was already proposed by Johann Wolfgang von Goethe in his ‘Farbenlehre’ (theory of colors; Goethe, 1867). In fact, many studies that obtained ratings on semantic differential scales to measure associations between color and meaning identified the warm–cool dimension as one important factor (Gao and Xin, 2006; Gao *et al.*, 2007; Kobayashi, 1981; Ou *et al.*, 2004; Sato *et al.*, 2000; Wright, 1962; Wright and Rainwater, 1962). In these studies, hues from the blue range were typically rated as cooler than hues from the red or yellow range, compatible with the assumption by Franz Marc. Besides hue, the saturation (chroma) and the brightness of a color (Wyszecki and Stiles, 2000) were reported to contribute to the ‘color warmth’. In some studies, highly saturated colors were perceived as warmer than colors with low saturations (Gao and Xin, 2006; Ou *et al.*, 2004; Wright and Rainwater, 1962; Xin *et al.*, 2004), while Wright (1962) found the color warmth to be mainly determined by the hue. Darker colors were sometimes also reported to be perceived as warmer than light colors (Wright and Rainwater, 1962). Note that Franz Marc did not consider the color dimensions brightness and saturation but talked only about the hue, corresponding to the hue-dominated concept of color used in everyday language.

A second meaning dimension mentioned by Marc is gender. He assumed blue to be masculine and yellow to be feminine. In fact, besides ‘color warmth’, factor analyses of semantic differential ratings for colors often identified a factor like potency (Osgood *et al.*, 1957) or ‘color weight’ (Wright, 1962), which in a study by Ou *et al.* (2004) included the semantic differential masculine–feminine, but also heavy–light and hard–soft. This factor depended mainly on the brightness, but also on the hue. However, in Ou *et al.*’s model the ‘color weight’ is lower for blue than for yellow and red. This is compatible with Marc’s assumption that red should be perceived as ‘heavy’, but not with his assumption that blue is perceived as ‘masculine’. However, because Ou *et al.* (2004) do not report the relation between the single scales for masculine–feminine and heavy–light and the hue, this apparent contradiction to Marc’s ideas could be due to the combination of the masculine–feminine and heavy–light scales into a single factor. More direct evidence is provided by Taft (1997) who obtained ratings on a masculine–feminine semantic differential scale for color photographs of objects like a sofa or a bicycle. Blue objects received ratings on the masculine side of the bipolar scale, while yellow and red received ratings on the feminine side of the scale.

As a third aspect, Franz Marc proposed ‘gentle’ and ‘brutal’ as associations with yellow and red, respectively. We are not aware of studies that obtained ratings for colors on the semantic differential brutal–gentle.

In our study, we focused on the meaning dimensions discussed above: gender (masculine–feminine), aggression (gentle–brutal), and temperature (warm–cool). The two former dimensions were selected because Franz Marc had proposed opponent color pairs for these opponent pairs of meaning categories (yellow–gentle/red–brutal and yellow–feminine/blue–masculine). We added the association blue–cool proposed by Marc and contrasted it with the association red–warm that was not explicitly mentioned by Marc, so that our experiment included the color pairs yellow–red, yellow–blue, and blue–red. In order to limit the duration of the experiment, we did not consider the remaining meaning dimensions implied by Marc's use of the adjectives stern, spiritual, cheerful, sensual, heavy, serious, and passive.

1.2. Implicit Testing Procedures

The aim of the present study was to evaluate certain hue–meaning associations proposed by Franz Marc. He developed his theory based on subjective perceptions and sensations; he constantly tried to "[…] organize color and make it the tool of artistic expression" (Marc and Macke, 1964, p. 45). Being an expressionist, he intended color to evoke an immediate emotional response within the viewer. For this reason, we were interested in whether the stated color–meaning associations proposed by Marc show up in an implicit testing procedure where participants are not required to explicitly express their opinion about the relation between a color and a semantic concept, as for example in the studies using semantic differentials discussed above. We used the Implicit Association Task (IAT) proposed by Greenwald *et al.* (1998), which is one of the implicit testing methods used most frequently (Fazio and Olson, 2003), and has been applied in the area of art and perception (Makin and Wuerger, 2013). In IATs, participants use two designated response buttons (e.g., left and right) for classifying stimuli into categories. For instance, in a study by Greenwald *et al.* (1998), participants were required to classify names of flowers and insects as either flowers or insects (termed targets or objects in the IAT literature), and to classify word stimuli (e.g., 'love', 'hatred'), termed attributes in the IAT literature, as being either positive or negative. Both flower/insect names and positive/negative words were presented in an experimental block, and the same response buttons were used for the classification as flower or insect and positive or negative. The fundamental assumption is that participants' responses are faster when associated and therefore congruent concepts share the same response button, compared to when non-associated (so-called incongruent) concepts share the same button. If one assumes, e.g., an association between flowers and pleasant words on the one hand and insects and unpleasant words on the other hand, response times (RTs) in the IAT should be faster when the concepts 'flower' and 'pleasant' are mapped to one response button and 'insect' and 'unpleasant' to the other button, compared to when the mapping is 'flower'–'unpleasant' for button 1 and 'insect'–'pleasant' for button 2. In fact, the IAT applied by Greenwald *et al.* (1998) was sensitive to evaluative differences: Congruent mappings resulted in a more rapid performance than incongruent pairings.

For the present purposes, a difficulty arises because the classic IAT described above requires complementary concepts (Greenwald and Farnham, 2000). While in Franz Marc's assumptions concepts like 'masculine–feminine' or 'cool–warm' are organized in opponent pairs, it is not possible to identify an unequivocal opposite of, for example, the hue-category 'red'. For this reason, we used the Single Category Implicit Association Test (SC-IAT) proposed by Karpinski and Steinman (2006). This method still includes two opponent attribute categories but, in contrary to the classic IAT, only one target category. To give an example, in a study by Bluemke and Friese (2008), a political party (e.g., the German CDU), shared the same response button with either the attribute 'positive' or the attribute 'negative'. In the experiment, subjects were asked to classify nouns (e.g., 'poison' or 'gift') as being either positive or negative, and to press the response button labeled 'CDU' when a logo of this political party or the name of a politician belonging to this party was presented on the screen. Again, response times are expected to be faster when the response mapping (e.g., 'CDU' and 'positive' sharing the same response button) matches the participant's attitude, compared to when the mapping is reversed (e.g., 'CDU' and 'negative' sharing the same button).

In our experiment, three different opponent pairs of meanings proposed by Franz Marc (e.g., 'masculine'–'feminine') were combined with the three hues (blue, red, yellow) in separate SC-IATs. In concordance with the corresponding color–meaning associations according to Franz Marc, we expected stronger associations between masculine–blue and feminine–yellow than between feminine–blue and masculine–yellow. For this reason, we expected shorter RTs for the two former than for the two latter response pairings. Note that Marc did not propose an association between red and either masculine or feminine. Following the same rationale, we expected faster RTs in the condition pairing cool–blue than for the pairing of warm–blue. Here, Marc did not propose an association between yellow or red and either cool or warm. Finally, the combination of gentle–yellow and brutal–red should evoke faster response times than the pairings of brutal–yellow and gentle–red, while Marc's theory does not predict an association between blue and either brutal or gentle.

2. Method

2.1. Participants

Thirty-six volunteers (20 females and 16 males) took part in our experiment. Their age ranged from 18 to 53 years with a mean age of $M = 26.7$ years

(SD = 8.1 years). All had normal or corrected-to-normal vision and normal color vision according to a short version of the Ishihara Color Test (Ishihara, 2013) using six plates. Twenty-nine of them were students or trainees and seven were working full-time. Psychology students received partial course credit for their participation. The other participants were not rewarded. The experiment was conducted in accordance with the principles stated in the Declaration of Helsinki. Prior to the experiment, every subject gave informed consent about the participation in a written form. The participants were not informed about the hypotheses under test.

2.2. *Stimuli and task*

We tested the color–meaning associations using SC-IATs (Karpinski and Steinman, 2006), measuring the association between one color category (e.g., red) and two meaning categories (e.g., masculine and feminine). In the IAT nomenclature, the color category served as target concept and the meaning categories as attribute dimensions. Throughout the experiment, participants had to classify different colors (target stimuli) and words (attribute stimuli) and assign them to the matching category by pressing the assigned response button. Color patches served as stimuli for the target concept ‘color’, and word stimuli for the meaning dimension (e.g., ‘masculine’ versus ‘feminine’).

As shown in table 1, we presented five word stimuli (‘attributes’) for each of the meaning categories (e.g., for ‘feminine’), as recommended by Greenwald *et al.* (1998) and Nosek *et al.* (2005). For the gender dimension (masculine vs. feminine), we used a word list proposed by Nosek *et al.* (2005). For the temperature dimension (warm vs. cool) and the aggression dimension (gentle vs. brutal), we selected adjectives from the German Duden dictionary, intending to find universal synonyms and antonyms. In order to prevent word length effects, the bipolar word pairs shared approximately the same number of letters, i.e., an equal number or a difference of plus or minus one letter, the only exception being: Junge (Boy) vs. Mädchen (Girl). For an IAT, the most important aspect is that each of the word stimuli clearly belongs to one of the two poles of the attribute dimension (e.g., masculine vs. feminine) and is therefore easy to classify. The word stimuli measured 1 cm in height (visual angle 1.0°) and, depending on the word, 4–8 cm in width (visual angle 3.8–7.6°).

For the color categories (‘targets’), we presented five different color patches for each hue (blue, yellow, red), resulting in 15 color stimuli. Their dimensions on the screen were 4 cm × 10 cm (visual angle 3.8° × 9.5°). Again, each of the color patches clearly represented one of the three hues, but within each hue category the colors differed in their exact hue as well as in brightness and saturation. The CIE 1931 *xyY* coordinates (Commission Internationale de l’Éclairage, 2006) of the color stimuli are shown in table 2.

Table 1.
Words (attribute stimuli) presented in the three SC-IATs (English translations in parentheses). The two opponent words shown in the first row for each SC-IAT are taken from Franz Marc's list of color–meaning associations (except 'warm', which was not explicitly listed by Marc). The words in the remaining rows represent the two opponent categories shown in the first row.

Gender SC-IAT		Temperature SC-IAT		Aggression SC-IAT	
Männlich	Weiblich	Warm	Kühl	Sanft	Brutal
(Masculine)	(Feminine)	(Warm)	(Cool)	(Gentle)	(Brutal)
Mann	Frau	Sonnig	Frostig	Sacht	Grob
(Man)	(Woman)	(Sunny)	(Frosty)	(Soft)	(Rough)
Junge	Mädchen	Sommerlich	Winterlich	Zart	Derb
(Boy)	(Girl)	(Summer)	(Winter)	(Tender)	(Coarse)
Er	Sie	Feurig	Eisig	Zärtlich	Grausam
(He)	(She)	(Fiery)	(Icy)	(Sensitive)	(Cruel)
Herr	Dame	Heiß	Kalt	Wohlwollend	Gewalttätig
(Sir)	(Lady)	(Hot)	(Cold)	(Benevolent)	(Violent)

Table 2.
CIE 1931 x–y coordinates and luminance (Y) of the color stimuli.

Color	x	y	Y (cd/m^2)
Blue 1	0.198	0.203	159.154
Blue 2	0.173	0.125	21.541
Blue 3	0.151	0.066	8.154
Blue 4	0.175	0.151	9.420
Blue 5	0.152	0.069	3.722
Yellow 1	0.394	0.492	102.468
Yellow 2	0.404	0.512	103.276
Yellow 3	0.397	0.449	85.001
Yellow 4	0.418	0.489	83.735
Yellow 5	0.436	0.487	76.727
Red 1	0.529	0.348	31.918
Red 2	0.596	0.328	21.079
Red 3	0.630	0.331	19.579
Red 4	0.630	0.331	22.664
Red 5	0.610	0.326	6.248

The mapping of the meaning categories and the color category to the two response buttons was shown by 'category reminders' in the two upper corners of the screen. As shown in Fig. 1, the two opponent meaning categories (e.g., masculine vs. feminine) were shown in the left and right upper corner and the single color category (e.g., red) was shown next to either the left or right attribute, in the form of a color word separated from the meaning category by an

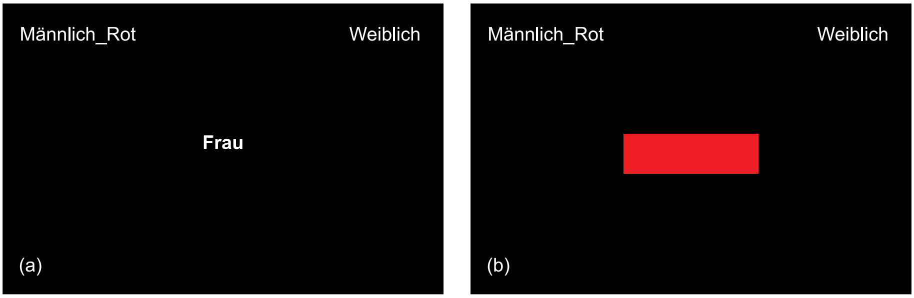

Figure 1. Gender SC-IAT. Example screenshots of the display with the categories ‘masculine’ and ‘red’ (männlich–rot) mapped to the left response button and the category ‘feminine’ (weiblich) mapped to the right response button. (a) Word stimulus: ‘Frau’ (‘woman’). The participant should press the right button. (b) Color stimulus: red. The participant should press the left button. Note the ‘category reminders’ in the upper corners showing the mapping of the two meaning categories and the color category to the two response buttons.

underscore. The category reminders measured 1.5 cm in height (visual angle 1.4°) and up to 8 cm in width (visual angle ≤ 7.6°).

The task was to classify color and word stimuli as belonging to particular categories. In order to do so, participants were instructed to position their left index finger on key ‘E’ (left key) and their right index finger on key ‘P’ (right key) on a German standard QWERTZ keyboard. Figure 1 shows two example screenshots of the displays with the categories ‘masculine’ and ‘red’ both assigned to the left button and the category ‘feminine’ assigned to the right button. In the left panel of Fig. 1, the word stimulus ‘Frau’ (‘woman’) has to be categorized as ‘feminine’ by pressing the right key. In the right panel, the color patch has to be categorized as ‘red’ by pressing the left key. Participants were asked to respond as quickly and as accurately as possible. The stimuli remained on the screen until the participant responded by pressing either the left or the right response key. We did not impose a response deadline (Cunningham *et al.*, 2001). If the response was correct, the next stimulus was presented after an inter-stimulus interval of 300 ms. If the response was incorrect, a large red cross (‘X’) appeared under the stimulus in the center of the screen. As soon as the participant entered the correct answer, the cross disappeared and the next stimulus was presented after a delay of 300 ms.

2.3. Procedure

The experimental blocks constituting each SC-IAT were designed according to the recommendations of Karpinski and Steinman (2006), but we also included a supplementary practice block for the word stimuli at the beginning of each SC-IAT, as in Bluemke and Friese (2008). In the latter block, the ten words selected for the relevant meaning dimensions (five word stimuli per meaning category; e.g. 'Lady' for the feminine category or 'Sir' for the masculine category in the Gender SC-IAT) were presented without pairing to a color. This block enabled participants to become familiar with the word stimuli and the response mapping. Because in each experimental block only one color category was presented, the color-categorization task was very simple so that we did not include any color-categorization practice trials.

Following the attribute discrimination task presented in the first block (word practice trials), participants completed six combined blocks, each consisting of 36 training trials [ten color trials (e.g., blue) mapped to one response button; ten word classification trials of the meaning category paired with the color (e.g., masculine) mapped to the same button as the color; 16 word classification trials of the unpaired meaning category (e.g., feminine) mapped to the other button], followed by 70 testing trials [20 color trials (e.g., blue) mapped to one response button; 20 word classification trials of the meaning category paired with the color (e.g., masculine) mapped to the same button as the color; 30 word classification trials of the unpaired meaning category (e.g., feminine) mapped to the other button]. The stimuli and the task in training and test trials were identical, but only the test trials were included in the data analysis. Note that as recommended by Bluemke and Friese (2008) and Karpinski and Steinman (2006), we presented slightly different proportions of left-hand and right-hand responses in the attempt to reduce response bias. During the test trials, 40 stimuli were assigned to one key (20 color patches and 20 words of the paired meaning category) and 30 stimuli to the other key (30 words of the unpaired meaning category) meaning that 57% of correct answers were allocated to one key and 43% to the other key. During the training trials, 20 of the 36 stimuli corresponded to one response button (ten color patches and ten words of the paired meaning category), and the remaining 16 to the other button (16 words of the unpaired meaning category). In order to present 16 word attributes for the unpaired meaning category, one word attribute had to be presented more frequently than the remaining words. For this purpose, we selected a word matching one of the category labels (masculine–feminine, warm–cool, gentle–brutal).

Each SC-IAT and each block were preceded by instructions explaining the categorization task, the appropriate key assignments and the following stimuli. All participants completed three SC-IATs in the same order: 1) Gender SC-IAT, 2) Temperature SC-IAT and 3) Aggression SC-IAT. We did not vary this order because the critical difference between the RTs in blocks with congruent and incongruent pairings of the color and the meaning category is not confounded by the order of meaning category pairs within the experiment. In IAT studies, fixed sequences of blocks are often used in order to minimize variability due to position effects (see Bluemke and Friese, 2008; Karpinski and Steinman, 2006; Penke *et al.*, 2006). Within each SC-IAT, there was a fixed mapping of the meaning categories to the left and right key (Gender SC-IAT: 'masculine' left, 'feminine' right; Temperature SC-IAT: 'warm' left, 'cool' right; Aggression SC-IAT: 'gentle' left, 'brutal' right).

Within each SC-IAT, we varied the sequence of blocks presenting the color categories blue, yellow, and red, resulting in six possible color sequences. In addition, for each color presented within a given SC-IAT (e.g., blue in the Aggression SC-IAT), we varied the sequence of pairings of the two meaning categories and the color, which results in two possible block orders. Either the color category was paired with the left meaning category (as in the example shown in Fig. 1) in the first block presenting a given color and therefore had to be categorized to the left side, followed by a block in which the color category was paired with the right meaning category, or vice versa. The six color sequences and two color position sequences were completely crossed, resulting in 12 possible orders of conditions. We randomly assigned three participants to each order of conditions, and the selected order of conditions was applied in all of the three SC-IATs presented to the participant. Table 3 shows an example sequence of experimental blocks. In the example, the color order was blue–yellow–red, and within the blocks presenting one hue, the color category was first assigned to the left key and then to the right key.

2.4. Apparatus

The experiment took place in a dimly lit booth in a laboratory space and took 45–60 minutes per participant. The stimuli were presented at an approximate distance of 60 cm on the center of a 17-inch CRT monitor (Dell M783p) with a resolution of 1024 × 768 pixels (horizontal × vertical), a refresh rate of 85 Hz and a color depth of 32 bit. Our computer possessed an Intel® Pentium® 4 CPU processor with 2.8 GHz, a 2.00 GB RAM main memory and an ATI Radeon X300/X550/X1050 series graphics card. We used the open source PsychoPy software for generating the SC-IATs (Peirce, 2007).

2.5. Data analysis

Each of the three SC-IATs was analyzed separately. The response time (RT) on correct trials was analyzed with repeated-measures analyses of variance (rmANOVAs) using a multivariate approach with Pillai's trace statistic. The average error rate was 7.0%. RTs shorter than 300 ms were excluded as anticipatory responses, which affected only 0.5% of the correct responses. Because the asymmetric distribution of RTs can cause problems in repeated-measures ANOVAs (e.g., Oberfeld and Franke, 2013), the RTs were log-transformed prior to all analyses (Greenwald *et al.*, 1998), $RT_{log} = \log_{10}(RT)$, where RT is the response time in milliseconds. For each combination of subject, SC-IAT (gender, temperature, aggression), color (blue, yellow, red), and response

Table 3.
Example of the sequence of experimental blocks.

SC-IAT	Block	Description	Left categories	Right categories	Number of trials
Gender	1	Attribute discrimination practice block	Masculine	Feminine	10
	2	Training	Masculine–Blue	Feminine	36
		Test	Masculine–Blue	Feminine	70
	3	Training	Masculine	Feminine–Blue	36
		Test	Masculine	Feminine–Blue	70
	4	Training	Masculine–Yellow	Feminine	36
		Test	Masculine–Yellow	Feminine	70
	5	Training	Masculine	Feminine–Yellow	36
		Test	Masculine	Feminine–Yellow	70
	6	Training	Masculine–Red	Feminine	36
		Test	Masculine–Red	Feminine	70
	7	Training	Masculine	Feminine–Red	36
		Test	Masculine	Feminine–Red	70
Temperature	1	Attribute Discrimination	Warm	Cool	10
	2	Training	Warm–Blue	Cool	36
		Test	Warm–Blue	Cool	70
	3	Training	Warm	Cool–Blue	36
		Test	Warm	Cool–Blue	70
	4	Training	Warm–Yellow	Cool	36
		Test	Warm–Yellow	Cool	70
	5	Training	Warm	Cool–Yellow	36
		Test	Warm	Cool–Yellow	70
	6	Training	Warm–Red	Cool	36
		Test	Warm–Red	Cool	70
	7	Training	Warm	Cool–Red	36
		Test	Warm	Cool–Red	70
Aggression	1	Attribute Discrimination	Gentle	Brutal	10
	2	Training	Gentle–Blue	Brutal	36
		Test	Gentle–Blue	Brutal	70
	3	Training	Gentle	Brutal–Blue	36
		Test	Gentle	Brutal–Blue	70
	4	Training	Gentle–Yellow	Brutal	36
		Test	Gentle–Yellow	Brutal	70
	5	Training	Gentle	Brutal–Yellow	36
		Test	Gentle	Brutal–Yellow	70
	6	Training	Gentle–Red	Brutal	36
		Test	Gentle–Red	Brutal	70
	7	Training	Gentle	Brutal–Red	36
		Test	Gentle	Brutal–Red	70

mapping of color and meaning category (e.g., ‘blue’ paired with ‘masculine’ vs. ‘blue’ paired with ‘feminine’), log-correct RTs more than three times the interquartile range above the third or below the first quartile were excluded as outliers (Tukey, 1977). This affected only 0.8% of the trials.

The hue (blue, yellow, red) and the paired meaning categories presented in the respective IAT (masculine–feminine, warm–cool, or gentle–brutal) served as within-subjects factors. The six color sequences and the two color position sequences were included as between-subjects factors. An α-level of 0.05 was used. Effect sizes are reported as partial η^2 and, for the case of tests comparing two conditions, as Cohen’s (1988) standardized mean difference measure d_z. In addition, we calculated so-called difference-scores (*D*-scores) that are often used in the analysis of IATs (Greenwald *et al.*, 2003). As discussed by Greenwald *et al.* (2003), the *D*-score is similar to the effect size measure *d* proposed by Cohen (1988) because for each participant and condition the difference between the mean RTs is expressed relative to the standard deviation of the RTs.

3. Results

3.1. Gender SC-IAT

As shown in Fig. 2, color–meaning response mappings that are congruent according to Franz Marc’s assumptions (i.e., blue–masculine and yellow–feminine) resulted in faster response times compared to incongruent response mappings (i.e., blue–feminine and yellow–masculine). Response times in blocks including the color red did not differ strongly between the pairing with either masculine or feminine. In the rmANOVA, the color (blue, yellow, red) × paired meaning category (feminine, masculine) interaction just failed to reach significance, $F(2, 23) = 3.38, p = 0.052, \eta_p^2 = 0.227$. We conducted three post-hoc ANOVAs, each comparing two colors. To control for the familywise error rate, we used a sequentially rejective step-up Bonferroni procedure (Hochberg, 1988). At an α-level of 0.05, the color × paired meaning category interaction was only significant in the analysis comparing blue and yellow, $F(1, 24) = 7.015$, $p = 0.014$, Cohen’s (1988) $d_z = 0.44$, corresponding to Franz Marc’s assumption that blue and yellow show opponent associations with masculine and feminine. As a second type of post-hoc tests, we computed paired-samples *t*-tests contrasting the log RTs for the pairing of each color with the two meaning categories. For blue, the log RTs were not significantly longer when the color was paired with feminine rather than masculine, $t(35) = 1.76$, $p = 0.087$, $d_z = 0.29$ (Cohen, 1988), although the descriptive trend (see Fig. 2) is compatible with the blue–masculine association proposed by Marc. For yellow, the log RTs were not significantly longer when the color was paired with masculine rather than feminine, $t(35) = 1.98$, $p = 0.056$, $d_z = 0.33$, while the

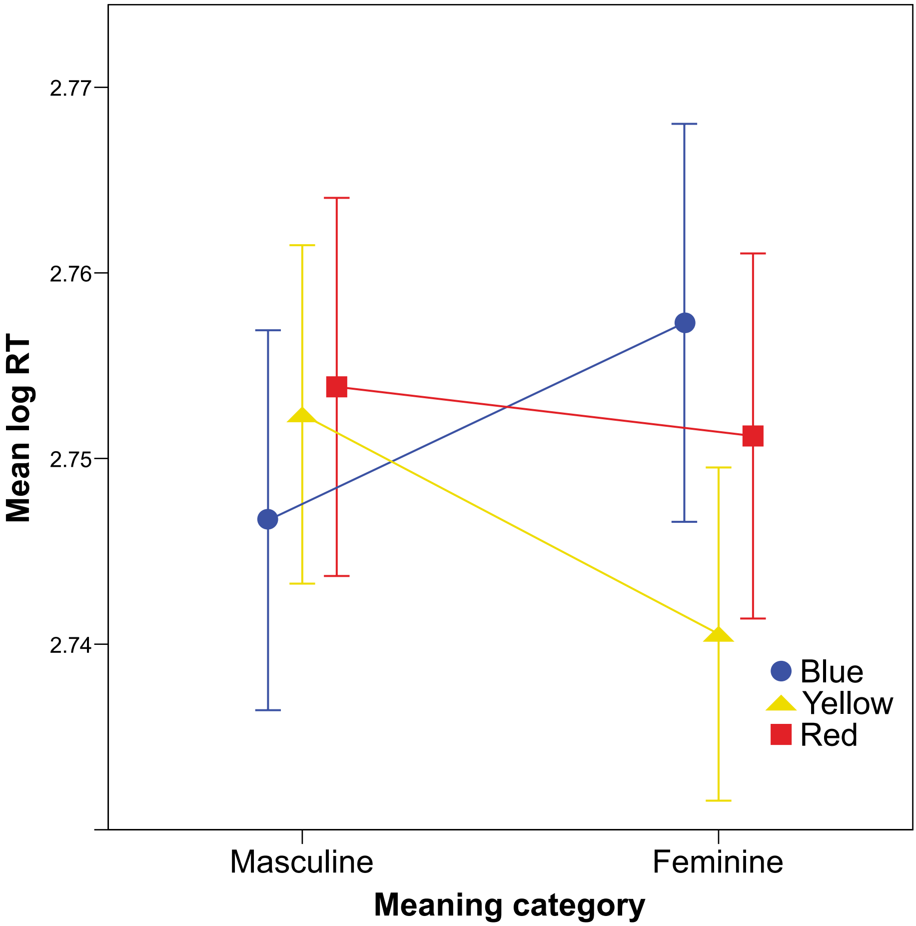

Figure 2. Mean log RTs in the Gender SC-IAT, as a function of the pairing between meaning category and color. Circles: blue; triangles: yellow; boxes: red. Error bars show plus and minus one standard error of the mean (SEM).

descriptive trend was again compatible with Marc's assumptions. For red, the log RTs did not differ significantly between the pairing of the color with masculine or feminine, $t(35) = 0.39$, $p = 0.70$, $d_z = 0.07$.

In the rmANOVA including all of the three colors, neither the effect of color, $F(2, 23) = 1.07$, $p = 0.359$, $\eta_p^2 = 0.085$, nor the effect of paired meaning category, $F(1, 24) = 0.14$, $p = 0.708$, $\eta_p^2 = 0.006$, was significant. The rmANOVA showed a significant between-subjects effect of color position sequence, $F(1, 24) = 6.86$, $p = 0.015$, $\eta_p^2 = 222$, $d_z = 0.437$. On average, participants who started with a block in which 'masculine' and the hue shared a response button responded faster than participants who started with blocks in which ‘feminine’ and the hue were mapped to the same button first (mean log RTs: $M = 2.73$, SD = 0.059, and $M = 2.77$, SD = 0.042, respectively). We have no explanation for this group difference. All remaining effects were not significant (all p-values > 0.114).

Figure 3 shows the mean *D*-scores (Greenwald *et al.*, 2003) for associations between color (blue, yellow, red) and meaning category (masculine, feminine). For each color, the *D*-scores were computed as the individual difference between the mean log RT in blocks where the color was paired with ‘feminine’ and the mean log RT in blocks where the color was paired with ‘masculine’, divided by the SD of the log RTs in all blocks of the Gender SC-IAT presenting the respective color, e.g., $D_{\text{f-m\_blue}} = (M_{\text{feminine–blue}} - M_{\text{masculine–blue}})/SD_{\text{blue}}$, where $M_{\text{feminine–blue}}$ is the mean log RT in blocks of the Gender SC-IAT where ‘blue’ and ‘feminine’ shared the same response button, $M_{\text{masculine–blue}}$ is the mean log RT in blocks where ‘blue’ and ‘masculine’ were mapped to the same response button, and SD_{blue} is the standard deviation of all log RTs in blocks of the Gender SC-IAT presenting the color blue. Thus, the *D*-score contrasts the RTs between the mappings of blue with either masculine or feminine, as in the paired-samples *t*-tests reported above, but taking into account the individual variability of the RTs. Not surprisingly, the mean *D*-scores showed a similar pattern as the paired-samples *t*-tests reported above. For blue, the average *D*-score was positive, although not significantly different from 0 (as shown by the 95% confidence intervals in Fig. 3). A positive *D*-score represents longer RTs for blue paired with ‘feminine’ than for blue paired with ‘masculine’. Thus, the observed non-significant trend is compatible with the blue–masculine association proposed by Franz Marc. For yellow, the average *D*-score tended to be negative, again compatible with the yellow–feminine pairing proposed by Marc. For red, the average *D*-score was close to 0. In accordance with this finding, Franz Marc had proposed no pairing of red with either masculine of feminine.

3.2. Temperature SC-IAT

As shown in Fig. 4, this SC-IAT showed faster response times when blue was paired with cool rather than warm, consistent with the assumption of Franz Marc. Also, for red and yellow the opposite pattern was observed, compatible with our assumption that these hues are associated with warm rather than cool. This pattern was confirmed by the rmANOVA which showed a significant color × paired meaning category interaction, $F(2, 23) = 12.19, p < 0.001$, $\eta_p^2 = 0.515$. Three post-hoc ANOVAs, each comparing two colors, resulted in significant color × paired meaning category interactions when comparing the SC-IATs for blue and red ($d_z = 0.82$) as well as for blue and yellow ($d_z = 0.69$), but not for the comparison of red and yellow (Hochberg procedure, α-level of

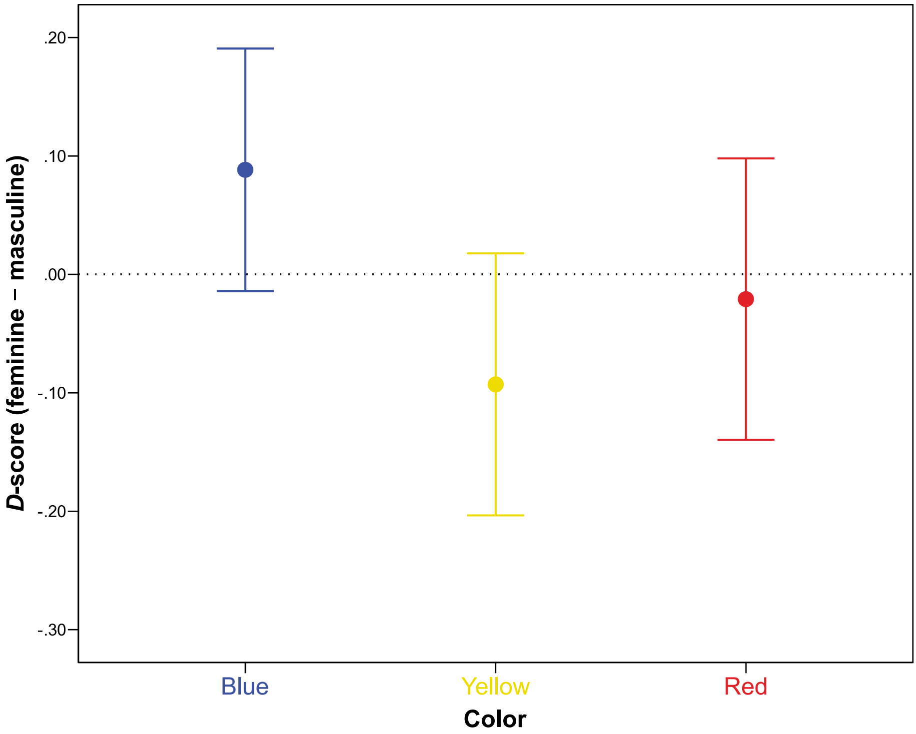

Figure 3. Mean D-scores for the Gender SC-IAT, as a function of color. Positive values represent an association of the color and the meaning category 'masculine' (opposed to 'feminine'). Error bars show 95% confidence intervals (CIs).

0.05). As a second type of post-hoc tests, we computed paired-samples *t*-tests contrasting the log RTs for the pairing of each color with the two meaning categories. For all colors, the difference in log RT between the two paired meaning categories was significant according to the Hochberg (1988) procedure (blue: $t(35) = 2.41$, $p = 0.022$, $d_z = 0.40$; yellow: $t(35) = 3.64$, $p = 0.001$, $d_z = 0.61$; red: $t(35) = 4.71$, $p < 0.001$, $d_z = 0.78$), compatible with Franz Marc's proposal that blue is associated with 'cool' and our additional assumption that red and yellow should be associated with 'warm'.

There was no significant effect of color in the main rmANOVA, $F(2, 23) = 0$ 0.67, $p = 0.520$, $\eta_p^2 = 0.055$, but a significant effect of paired meaning category, $F(1, 24) = 13.00$, $p = 0.001$, $\eta_p^2 = 0.351$. A comparison of mean log RTs for the two different attributes paired with the color showed that participants responded faster in blocks where the meaning category 'warm' shared the response button with the color ($M = 2.77$, SD = 0.070) than in those were the category 'cool' and the color shared the response button ($M = 2.79$, SD = 0.070).

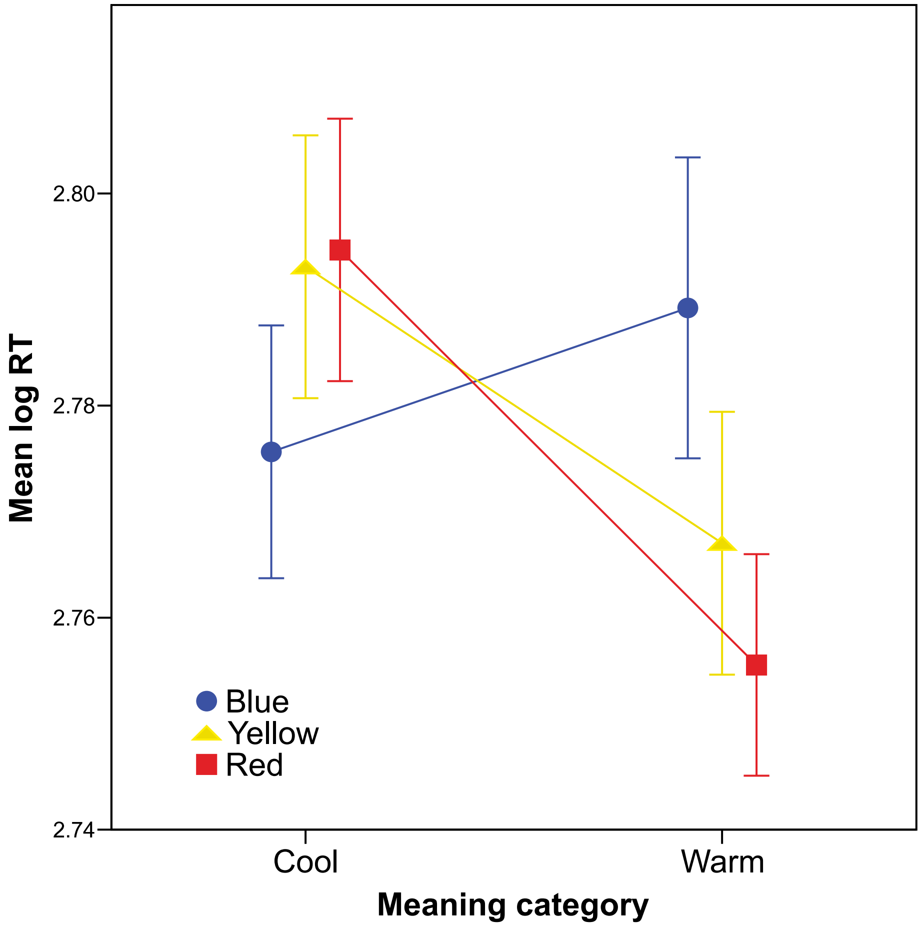

Figure 4. Mean log RTs in the Temperature SC-IAT, as a function of the pairing between meaning category and color. Circles: blue; triangles: yellow; boxes: red. Error bars show ±1 SEM.

We have no explanation for this effect. The effects of the between-subject factors and all other remaining effects were not significant (all p-values > 0.088).

The mean *D*-scores for associations between color (blue, yellow, red) and the meaning category 'warm' (as opposed to 'cool') are shown in Fig. 5. They were computed analogically to the *D*-scores of the Gender SC-IAT (e.g., $D_{\text{c–w\_blue}} = (M_{\text{cool–blue}} - M_{\text{warm–blue}})/SD_{\text{blue}}$). The average *D*-score for blue was negative, which is in accordance with Franz Marc's assumption of an association of blue and cool, but just failed to reach significance. The *D*-scores for yellow and red were positive and significantly different from 0, indicating longer RTs when these colors were paired with 'cool' compared to when they were paired with 'warm'. According to the *D*-scores and according to

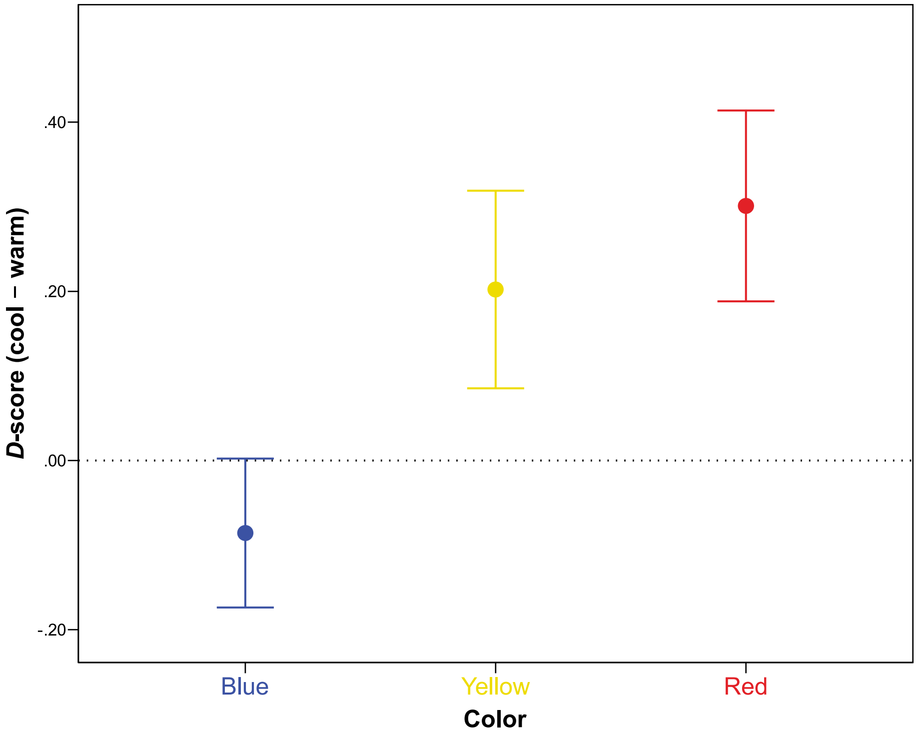

Figure 5. Mean D-scores for the Temperature SC-IAT, as a function of color. Positive values represent an association of the color and the meaning category 'warm' (opposed to 'cool'). Error bars show 95% CIs.

the effect sizes reported above for the paired-samples *t*-tests, the association of red–warm is even stronger than for yellow–warm. These two associations, however, were not mentioned by Franz Marc.

3.3. Aggression SC-IAT

In line with Marc's assumption, the congruent color–meaning response mapping of yellow–gentle in this SC-IAT resulted in faster RTs compared to the incongruent response mapping of yellow–brutal (see Fig. 6). However, the RTs were very similar for the congruent response mapping of red–brutal and the incongruent mapping of red–gentle. For blue, the RTs did also not differentiate strongly between the pairing with either brutal or gentle.

The rmANOVA showed a significant color × paired meaning category interaction, $F(2, 23) = 5.85$, $p = 0.009$, $\eta_p^2 = 0.337$. Thus, compatible with our expectations, the RTs depended on the pairing of color and meaning category. Separate post-hoc ANOVAs for each pair of colors indicated that the color × paired meaning category interaction was only significant when comparing the

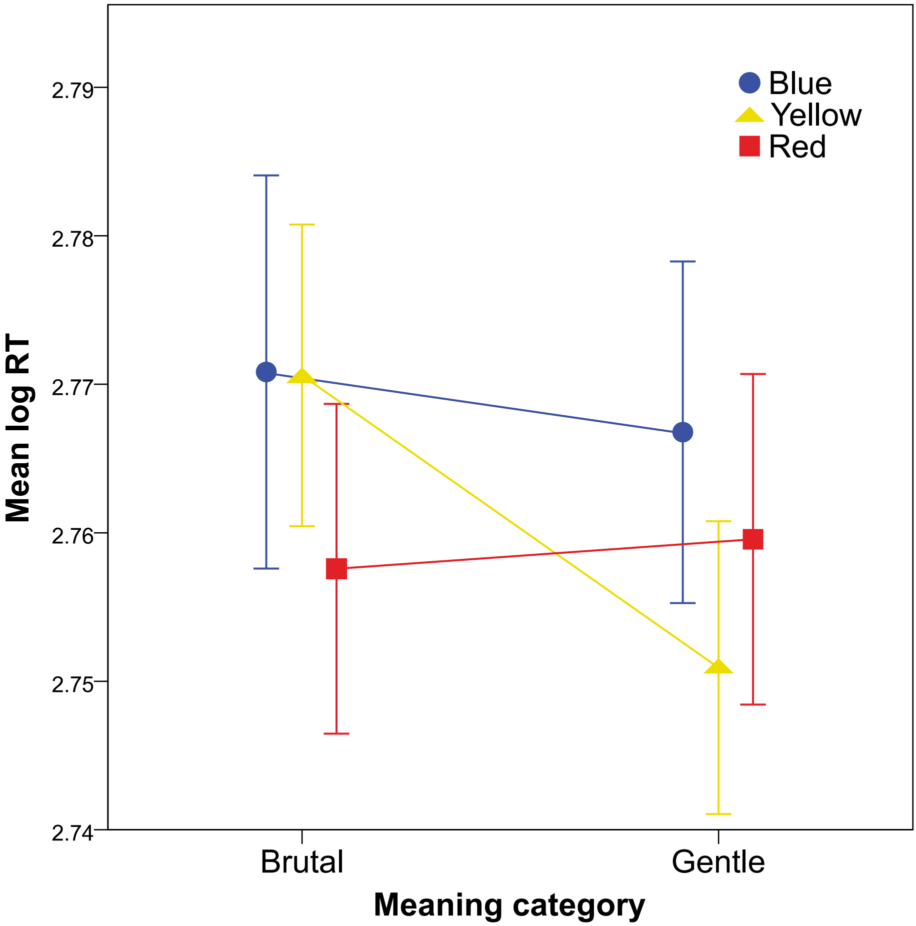

Figure 6. Mean log RTs in the Aggression SC-IAT, as a function of the pairing between meaning category and color. Circles: blue; triangles: yellow; boxes: red. Error bars show ±1 SEM.

SC-IATs for yellow and red (Hochberg procedure; $d_z = 0.57$). As a second type of post-hoc tests, we again computed paired-samples *t*-tests contrasting the log RTs for the pairing of each color with the two meaning categories. Only for yellow a significant difference in log RT between the two paired meaning categories was observed (blue: $t(35) = 0.55$, $p = 0.59$, $d_z = 0.09$; yellow: $t(35) = 4.81$, $p < 0.001$, $d_z = 0.80$; red: $t(35) = 0.34$, $p = 0.74$, $d_z = 0.06$), compatible with Franz Marc's proposal that yellow is associated with 'gentle', while his proposed association between red and 'brutal' was not confirmed.

In the main rmANOVA, there were no other significant effects (all p-values > 0.081).

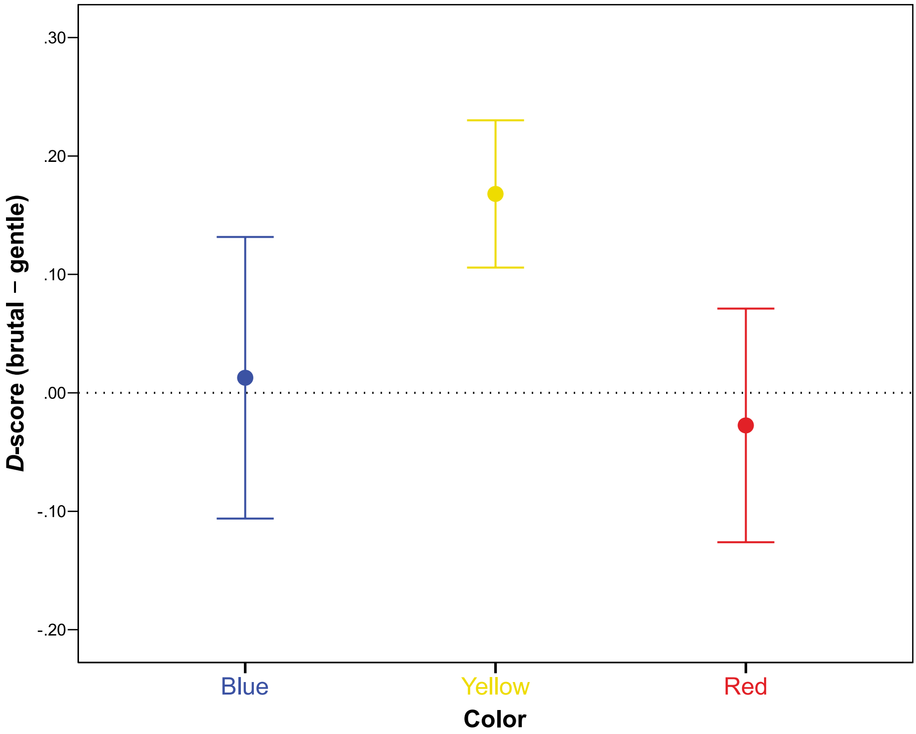

Figure 7. Mean D-scores for the Aggression SC-IAT, as a function of color. Positive values represent an association of the color and the meaning category 'gentle' (opposed to 'brutal'). Error bars show 95% CIs.

The average *D*-scores for the Aggression SC-IAT were computed for associations between color (blue, yellow, red) and meaning category (gentle, brutal), e.g., $D_{\text{b–g\_blue}} = (M_{\text{brutal–blue}} - M_{\text{gentle–blue}})/SD_{\text{blue}}$. As shown in Fig. 7, for yellow, a significantly positive *D*-score was observed, which means that yellow paired with brutal resulted in longer RTs then yellow paired with gentle. This speaks in favor of Marc's proposed association of yellow and gentle. The *D*-score for red was weakly negative, but not significantly different from 0. Thus, the RTs did not confirm the association of red and brutal proposed by Franz Marc. Marc did not mention any associations with blue in this context, which is in accordance with the *D*-score for blue being close to 0 and therefore not differentiating RTs in the pairing with either gentle or brutal.

4. Discussion

We tested a subset of the hue–meaning associations proposed by the expressionist painter Franz Marc (Der blaue Reiter) (Marc and Macke, 1964), using response time-based implicit association tests. In sum, certain assumptions of Franz Marc were confirmed by our results, others received no empirical support, and additional hue–meaning associations not proposed by Franz Marc were identified.

The first SC-IAT (Karpinski and Steinman, 2006) confirmed the associations blue–masculine and yellow–feminine proposed by Franz Marc. Response times were significantly shorter in blocks with these congruent response mappings, compared to blocks with incongruent mappings, as indicated by the color × paired meaning category interaction in the post-hoc rmANOVA of the log RTs for the colors blue and yellow. However, judging from the effect sizes, the associations between blue and yellow and either masculine or feminine were relatively weak, and less clear cut than, for instance, the associations between hue and temperature (see below). The response times did not indicate a significant association of red with either masculine or feminine, and in fact Franz Marc had not proposed such an association. As discussed in the introduction, a study by Taft (1997) measuring color–meaning associations had included a 'masculine'–'feminine' semantic differential, and reported that blue was associated with 'masculine', while red and yellow were associated with 'feminine'. Here, we assessed the associations between hue and the masculine–feminine dimension using an implicit, RT-based task. Although it is not possible to perform a quantitative comparison between the color–gender associations observed in Taft (1997) and in our study, the fact that our RT-based measures indicated a comparable pattern of (weak) associations between blue/yellow and feminine/masculine speaks to the robustness of these color–meaning correspondences. For red, the rating data by Taft (1997) showed an association with feminine rather than with masculine, while our implicit measure did not indicate an association of red with either gender.

In the second SC-IAT, we tested the association of blue and cool proposed by Franz Marc, and the associations between red and yellow and warm, which we had expected based on previous results concerning 'color warmth' (e.g., Wright and Rainwater, 1962). Response times were significantly shorter in the congruent condition blue–cool than in the incongruent condition blue–warm, although the *D*-score was not significantly different from 0. Thus, the association between hue and temperature as judged from the *D*-scores was stronger for red and yellow, for which Franz Marc had not proposed an association with either warm or cool. However, our results are compatible with previous data showing that 'color warmth' depends mainly on the hue, with blue hues being associated with cool and yellow or red hues being associated with warm (Kobayashi, 1981; Ou *et al.*, 2004; Sato *et al.*, 2000; Taft, 1997; Wright, 1962; Wright and Rainwater, 1962; Xin *et al.*, 2004). All of the latter studies asked for explicit ratings of associations between a color and temperature. Our data represent a complement to existing results because we assessed the associations between hue and the warm–cool dimension using an implicit, RT-based task. How can the color–temperature associations be explained? A role of changes in the daylight spectrum has been proposed (e.g., Huebner *et al.*, 2016), although the association with colors of the environment and temperature are not always clear cut. For example, on a day with clear sky, the sky is blue at noon but reddish or yellowish at dusk or dawn. However, on average, the ambient temperature is *higher* at noon than at dusk or dawn, so that the association blue–cool cannot be explained on this basis. A pattern compatible with the blue–cool association arises in a situation with direct sunlight: in the sun, colors tend to be yellowish while in the shadow they tend to be bluish. It should be noted that changes in the daylight spectrum or in color statistics of natural scenes across day times or seasons are relatively complex (Hernandez-Andres *et al.*, 2001; Lee and Hernandez-Andres, 2005; Webster *et al.*, 2007), and that humans are not highly sensitive in judging the time of day, month, or temperature from pictures of natural scenes (Granzier and Valsecchi, 2014). It appears that color–temperature associations of specific objects are more clear cut, like a warm, red/yellow bonfire or cool blue water. On a more general level, our data do not permit to decide whether the color–meaning associations are 'innate' as opposed to 'cultural'/'learned'.

Comparing the overall mean RTs of all three SC-IATs, the Temperature SC-IAT resulted in the longest RTs. This observation is accompanied by many comments of our participants stating that the second SC-IAT was the most difficult, presumably because the classification of the word stimuli as 'warm' or 'cool' (see Table 1) was more difficult than the classifications in the other two SC-IATs.

The association between red and 'brutal' as well as yellow and 'gentle' proposed by Franz Marc was tested in the third SC-IAT. The RTs and the *D*-scores were compatible with an association of yellow and gentle, but did not show clear associations with red and blue and either brutal or gentle. We are not aware of previous studies assessing associations between color and the brutal–gentle dimension with either explicit or implicit measures. However, a recent study (Dael *et al.*, 2017) measured the associations between color words and discrete emotions using the Geneva Emotion Wheel (Scherer, 2005). The emotion 'anger', which has an aggressive connotation, was relatively strongly associated with red in participants from Germany and the UK, and to a limited extent also in participants from China and Greece. However, red was also associated with the positive (and non-aggressive) emotion 'love' in all of the four countries (see also Sutton and Altarriba, 2016). This pattern is compatible with the lack of a clear association of red with either 'gentle' or 'brutal' observed in our experiment. For yellow, Dael *et al.* (2017) found an association with the 'aggressive' emotion 'hate' only in participants from Greece, while in all of the four countries yellow was associated with positive emotions like joy, pleasure, or contentment. This is again compatible with the significant yellow–gentle association observed here. However, blue was associated only with discrete emotions from the positive spectrum in Dael *et al.* (2017), while the Aggression SC-IAT showed no clear association between blue and either 'gentle' or 'brutal'. Taken together, the aggression dimension in relation to color is less well researched than for example the temperature dimension.

The implicit testing procedure (SC-IAT; Karpinski and Steinman, 2006) used in the present experiment measures 'automatic' associations (Greenwald *et al.*, 2003). Compared to an explicit assessment of color–meaning associations, for example by asking the subject to rate a color on a masculine–feminine semantic differential scale, this measure avoids potential effects of social desirability, because the participants are typically not aware of the associations the experimenter intends to test. In addition, the effect of a mapping of color and meaning on the response time in a simple classification task shows that the color–meaning association is a relatively fast process that does not require conscious deliberation. For this reason, the fact that our study confirmed several color–meaning associations with an implicit testing procedure speaks to the robustness and validity of these associations. As a cautionary note, one should not, however, draw the conclusion that an effect of color–meaning mappings in an RT-based task shows that these associations are 'low level' in the sense that they do not involve cognition.

Franz Marc proposed associations between German meaning categories and hue, and our experiment tested these associations using German words. The color–meaning associations might be different in other languages and other cultures, and it remains for future research to study such associations with implicit testing procedures in different languages. Previous cross-cultural studies on color–meaning and color–emotion associations showed both similarities and differences between cultures and languages (e.g., Adams and Osgood, 1973; Dael *et al.*, 2017; D'Andrade and Egan, 1974; Gao, *et al.*, 2007; Hupka *et al.*, 1997). For example, in the study by Dael *et al.* (2017), participants from China, Germany, Greece, and the UK consistently associated only positive emotions with the color pink but only negative emotions with the color gray. However, there were also a few systematic differences. For example, the color white was associated with sadness in China, but not in the other three countries.

When telling his friend August Macke about his ideas of colors and their association, Franz Marc was talking about his own individual sensation. This study confirms parts of his 'color theory' empirically. It would be interesting to test the additional color–meaning associations mentioned by Franz Marc, for instance blue–spiritual and red–heavy.

In the SC-IATs, the color category reminders present on top of the screen (see Fig. 1) were presented as color words printed in an achromatic color, rather than as, for example, patches with a chromatic color. It appears likely that the 'images' of color elicited by the color terms 'blue', 'red', and 'yellow' were relatively 'typical' exemplars with a high saturation, rather than some lowly saturated variants. At this point, it is important to recall that in his brief formulation of a 'color theory' Franz Marc considered only the hue, but not the brightness and saturation of colors. However, from other parts of his correspondence with August Macke, it is evident that he was well aware that there are for example many different types of blue, and that these types of blue can have different functions in a painting or can carry different meanings. In fact, previous research has identified effects of brightness and saturation on emotion (e.g., D'Andrade and Egan, 1974; Valdez and Mehrabian, 1994; Wilms and Oberfeld, 2017) as well as effects of these color dimensions on color warmth or color weight (Ou *et al.*, 2004; Wright and Rainwater, 1962). It remains for future research to investigate meaning associations for colors varying not only in their hue but also in saturation and brightness using implicit testing procedures.

To summarize, our data show that some associations between color and meaning are confirmed in implicit tests, and that bringing together the heuristic science of art and the empirical science of psychology opens up interesting topics in the fields of perception and emotion.

Acknowledgements

We gratefully thank Agnes Münch for her support with the programming of the Single Category Implicit Association Tests and Marius Frenken for his assistance in organizing the references and preparing the figures.

Note

1. Original German text: "Blau ist das männliche Prinzip, herb und geistig. Gelb das weibliche Prinzip, sanft, heiter und sinnlich. Rot die Materie, brutal und schwer und stets die Farbe, die von den anderen beiden bekämpft und überwunden werden muss! Mischst Du z. B. das ernste, geistige Blau mit Rot, dann steigerst Du das Blau bis zur unerträglichen Trauer, und das versöhnende Gelb, die Komplementärfarbe zu Violett, wird unerlässlich… Mischst Du Rot und Gelb zu Orange, so gibst Du dem passiven und weiblichen Gelb eine ‚megärenhafte', sinnliche Gewalt, dass das kühle, geistige Blau wiederum unerlässlich wird… Mischst Du nun aber Blau und Gelb zu Grün, so weckst du Rot, die Materie, die ‚Erde', zum Leben…".

References

Adams, F. M. and Osgood, C. E. (1973). A cross-cultural study of the affective meanings of color, *J. Cross-Cult. Psychol.* **4**, 135–156.

Bluemke, M. and Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects, *Eur. J. Soc. Psychol.* **38**, 977–997.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. L. Erlbaum Associates, Hillsdale, NJ, USA.

Commission Internationale de l'Éclairage. (2006). *Colorimetry—Part 1: CIE Standard Colorimetric Observers*. CIE S 014-1/E:2006 / ISO 11664-1:2007. URL: http://www.cie.co.at/publications/cie-colorimetry-part-1-standard-colorimetric-observers.

Cunningham, W. A., Preacher, K. J. and Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity, *Psychol. Sci.* **12**, 163–170.

Dael, N., Jonauskaite, D., Oberfeld, D., Zhang, M., Papadatou-Pastou, M. A., Havelka, J. and Mohr, C. (2017). *Common Factors Underlying Colour–Emotion Associations in China, Germany, Greece and United Kingdom*. Poster presented at the International Convention of Psychological Science, Wien, Austria. doi: 10.13140/RG.2.2.29787.87846.

D'Andrade, R. and Egan, M. (1974). The colors of emotion, *Am. Ethnol.* **1**, 49–63.

Fauth, D. (2010). *Kunstlexikon: Expressionismus*. URL: http://www.hatjecantz.de/expressionismus-5042-0.html. Accessed 22 December, 2017.

Fazio, R. H. and Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use, *Annu. Rev. Psychol.* **54**, 297–327.

Gao, X. P. and Xin, J. H. (2006). Investigation of human's emotional responses on colors, *Color Res. Appl.* **31**, 411–417.

Gao, X. P., Xin, J. H., Sato, T., Hansuebsai, A., Scalzo, M., Kajiwara, K., Guan, S.-S., Valldeperas, J., Lis, M. J. and Billger, M. (2007). Analysis of cross-cultural color emotion, *Color Res. Appl.* **32**, 223–229.

Goethe, J. W. von (1867). *Zur Farbenlehre. Didaktischer Theil.* (Band 33). Verlag der Cotta'schen Buchhandlung, Stuttgart, Germany.

Granzier, J. J. M. and Valsecchi, M. (2014). Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions, *J. Vis.* **14**. doi: 10.1167/14.1.22.

Greenwald, A. G. and Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept, *J. Pers. Soc. Psychol.* **79**, 1022–1038.

Greenwald, A. G., McGhee, D. E. and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test, *J. Pers. Soc. Psychol.* **74**, 1464–1480.

Greenwald, A. G., Nosek, B. A. and Banaji, M. R. (2003). Understanding and using the Implicit Association Test. I. An improved scoring algorithm, *J. Pers. Soc. Psychol.* **85**, 197–216.

Hernandez-Andres, J., Romero, J., Nieves, J. L. and Lee, R. L. (2001). Color and spectral analysis of daylight in southern Europe, *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **18**, 1325–1335.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika* **75**, 800–802.

Huebner, G. M., Shipworth, D. T., Gauthier, S., Witzel, C., Raynham, P. and Chan, W. (2016). Saving energy with light? Experimental studies assessing the impact of colour temperature on thermal comfort, *Energy Res. Soc. Sci.* **15**, 45–57.

Hupka, R. B., Zaleski, Z., Otto, J., Reidl, L. and Tarabrina, N. V. (1997). The colors of anger, envy, fear, and jealousy: A cross-cultural study, *J. Cross-Cult. Psychol.* **28**, 156–171.

Ishihara, S. (2013). *The Series of Plates Designed as a Test of Colour Deficiency. 24 Plates Edition.* Kanehara Trading Inc., Tokyo, Japan.

Karpinski, A. and Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition, *J. Pers. Soc. Psychol.* **91**, 16–32.

Kobayashi, S. (1981). The aim and method of the color image scale, *Color Res. Appl.,* **6**, 93–107.

Lee, R. L. and Hernandez-Andres, J. (2005). Short-term variability of overcast brightness, *Appl. Opt.* **44**, 5704–5711.

Makin, A. D. J. and Wuerger, S. M. (2013). The IAT shows no evidence for Kandinsky's color–shape associations, *Front. Psychol.* **4**. doi: 10.3389/fpsyg.2013.00616.

Marc, F. and Macke, A. (1964). *Briefwechsel 1910–1914*. DuMont, Köln, Germany.

Melcher, D. and Bacci, F. (2013). Perception of emotion in abstract artworks: A multidisciplinary approach, in: *The Fine Arts, Neurology, and Neuroscience: New Discoveries and Changing Landscapes,* S. Finger, D. W. Zaidel, F. Boller and J. Bogousslavsky (Eds), pp. 191–216. Elsevier, Oxford, UK.

Nosek, B. A., Greenwald, A. G. and Banaji, M. R. (2005). Understanding and using the Implicit Association Test. II. Method variables and construct validity, *Pers. Soc. Psychol. Bull.* **31**, 166–180.

Oberfeld, D. and Franke, T. (2013). Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data, *Behav. Res. Meth.* **45**, 792–812.

Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press, Urbana and Chigaco.

Ou, L. C., Luo, M. R., Woodcock, A. and Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Res. Appl.* **29**, 232–240.

Partsch, S. (2005). *Franz Marc 1880–1916*. Taschen, Köln, Germany.

Partsch, S. (2010). *Kunst Epochen 20. Jahrhundert I.* Reclam, Stuttgart, Germany.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python, *J. Neurosci. Meth.* **162**, 8–13.

Penke, L., Eichstaedt, J. and Asendorpf, J. B. (2006). Single-Attribute Implicit Association Tests (SA-IAT) for the assessment of unipolar constructs—The case of sociosexuality, *Exp. Psychol.* **53**, 283–291.

Sato, T., Kajiwara, K., Hoshino, H. and Nakamura, T. (2000). Quantitative evaluation and categorising of human emotion induced by colour, *Adv. Colour Sci. Technol.* **3**, 53–59.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci.Inf.* **44**, 695–729.

Sutton, T. M. and Altarriba, J. (2016). Color associations to emotion and emotion-laden words: A collection of norms for stimulus construction and selection, *Behav. Res. Meth.* **48**, 686–728.

Taft, C. (1997). Color meaning and context comparisons of semantic ratings of colors on samples and objects, *Color Res. Appl.* **22**, 40–50.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Pub. Co., Reading, MA, USA.

Valdez, P. and Mehrabian, A. (1994). Effects of color on emotions, *J. Exp. Psychol.* **123**, 394–409.

Van Paasschen, J., Zamboni, E., Bacci, F. and Melcher, D. (2014). Consistent emotions elicited by low-level visual features in abstract art, *Art Percept.* **2**, 99–118.

Vriesen, G. and Imdahl, M. (1967). *Robert Delaunay—Licht und Farbe*. DuMont, Köln, Germany.

Webster, M. A., Mizokami, Y. and Webster, S. M. (2007). Seasonal variations in the color statistics of natural images, *Network* **18**, 213–233.

Wilms, L. and Oberfeld, D. (2017). Color and emotion: Effects of hue, saturation, and brightness, *Psychol. Res*. 2017. doi: 10.1007/s00426-017-0880-8.

Wright, B. (1962). The influence of hue, lightness, and saturation on apparent warmth and weight, *Am. J. Psychol.* **75**, 232–241.

Wright, B. and Rainwater, L. (1962). The meanings of color, *J. Gen. Psychol.* **67**, 89–99.

Wyszecki, G. and Stiles, W. S. (2000). *Color Science: Concepts and Methods, Quantitative Data, and Formulae* (Wiley classics library ed.). John Wiley and Sons, New York, NY, USA.

Xin, J. H., Cheng, K. M., Taylor, G., Sato, T. and Hansuebsai, A. (2004). Cross-regional comparison of colour emotions. Part I: Quantitative analysis, *Color Res. Appl.* **29**, 451–457.