



Contents lists available at ScienceDirect

Food Quality and Preference

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

Using color–odor correspondences for fragrance packaging design

Hendrik N.J. Schifferstein^{a,*}, Bryan F. Howell^b^a Department of Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands^b Department of Industrial Design, School of Technology, Ira R. Fulton College of Engineering, Brigham Young University, 265 Crabtree Building, Provo, UT, USA

ARTICLE INFO

Article history:

Received 11 March 2015
 Received in revised form 1 June 2015
 Accepted 16 June 2015
 Available online 17 June 2015

Keywords:

Fragrance
 Color
 Congruency
 Package
 Design
 Buying intention

ABSTRACT

Fragrance companies may use colors in packaging design to communicate the properties of their fragrances. Packages with matching colors may have an additional advantage: some studies suggest that consumers prefer offerings for which all sensory impressions are congruent. Hence, we investigated whether consumers are more likely to buy fragrances with a matching package.

For five fragrances we created a package, based on a harmonious combination of three colors that obtained high odor–color goodness-of-fit ratings in a previous study. Packages were identical in design, except for the colors used. The appropriateness ratings for the five packages were indeed found to be related to the ratings for its major colors.

Subsequently, we assessed the degree to which participants were likely to buy different fragrance–package combinations. However, buying intentions were not affected by the degree of matching between fragrance and packaging. Instead, they were typically dependent on the degree of liking for the fragrance and, to a lesser extent, on the degree of liking for the package.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Communicating fragrance properties is difficult without directly experiencing the smell of the fragrance. Fortunately, several studies have shown that people show fairly consistent odor–color correspondences. Possibly, fragrance companies can make use of such cross-modal correspondences to create fragrance packages and advertisements that communicate the properties of these fragrances and thus generate consumer expectations that are as aligned as possible with the fragrance perception. Another reason why companies may make use of these correspondences is that consumers may prefer offerings for which all sensory impressions are congruent. If different sensory channels communicate corresponding messages, the product offering is likely to be perceived as more unitary, which could result in a higher degree of liking and appreciation, and higher purchase likelihood.

In the present paper, we create different multi-color fragrance packages based on the goodness-of-fit judgments for a number of specific fragrance–color combinations. We show that packages that are composed of three matching colors also match that particular fragrance when they are presented in the form of a multicolor

package. Subsequently, we investigate whether matching packages result in increased buying intentions for the target fragrances.

1.1. Odor–color correspondences

People have difficulty in communicating about smell experiences through words (Engen, 1982), which limits the usefulness of verbal communication in fragrance marketing. As a consequence, fragrance companies who want to inform potential consumers about the experiences their products elicit face a difficult task. An alternative could be to communicate properties of fragrance experiences through the colors of the fragrance fluid, the bottle, the packaging, advertisements, and other promotion materials (Scharf & Volkmer, 2000).

Several authors have reported consistent relationships between perceived odor properties and color properties. In one of the first systematic studies on the topic, Gilbert, Martin, and Kemp (1996) found that color matches for various equi-intense fragrance materials differed on the Munsell dimensions chroma (saturation) and value (brightness). In a follow-up study in which the odorants also varied in intensity, Kemp and Gilbert (1997) found that matching colors generally showed greater variation in hue across than within odors, while perceived intensity was related to brightness. Schifferstein and Tanudjaja (2004) performed an odor–color matching study in which they used complex fine fragrances and color chips from the NCS color system as stimuli. In this study,

* Corresponding author.

E-mail addresses: h.n.j.schifferstein@tudelft.nl (H.N.J. Schifferstein), bryan.howell@byu.edu (B.F. Howell).

the blackness dimension yielded the largest number of significant differences between fragrances, suggesting that the color brightness dimension may be most salient in distinguishing between fragrances. More recently, Kim (2013) showed that color dimensions can be related to the different types of fragrance families.

The analysis of these matching studies conducted with fragrances and their ingredient materials has revealed relationships between dimensions underlying odor perception (intensity, odor quality) and the dimensions underlying color perception (hue, saturation, brightness). However, these dimensional interrelationships are often very coarse, overly generalizing, and not specific enough to provide the exact fragrance or color properties designers would require for designing a new product. Hence, to obtain more detailed information Schifferstein and Tanudjaja (2004) also asked participants to rate the degree-of-fit between samples of 14 fine fragrances and 17 colors samples. These data showed consistent differences in associated colors between the different fragrances. For example, the fragrance Wish was mainly associated with Red and Orange, while Kouros was associated with Blue, and DKNY with Yellow. Maric and Jacquot (2013) expanded on this research by showing that participants were also able to make subtle discriminations between samples eliciting similar odor percepts (e.g., lime and lemon) by selecting different color matches (more Green and Brown for lime than for lemon). The existence of odor–color associations has not only been shown in self-report studies, but has also been confirmed in implicit association tasks for a number of specific odor–color pairs (Demattè, Sanabria, & Spence, 2006).

Cross-modal correspondences may stem from a perceptual, a semantic or an affective mechanism (e.g., Schifferstein & Tanudjaja, 2004; Stevenson, Rich, & Russell, 2012). Perceptual stimulus representations may exhibit similarity, because they result from similar neural patterns, even though these were generated by different sets of sensory receptors. For instance, perceptual magnitude (weak versus strong) may be encoded in similar ways for different sensory modalities. A semantic origin is likely when stimuli have acquired the same meaning through everyday experiences. For instance, when seeing the Brown color of cinnamon has often occurred together with the smell of cinnamon, people are likely to think of a Brown color when they smell cinnamon on a subsequent occasion. Even when people experience difficulty in identifying the smell of a banana, this odor may elicit the association with fruit or, even more general, with food. Thereby, these associations limit the range of colors that seem appropriate. In case of a hedonic origin of correspondences, people will associate stimuli that they like to the same degree. For instance, Schifferstein and Tanudjaja (2004) found that the degree-of-fit between odors and colors increased when stimulus ratings on the pleasure dimension became more similar.

1.2. Possible effects of (in)congruent package colors on odor experience

But what are the effects of a high odor–color congruency? Recently, Zellner (2013) reviewed the literature on odor–color correspondences and their implications for odor–color interactions. An appropriate color usually helps in identifying an odor (Davis, 1981; Zellner, Bartoli, & Eckard, 1991) and increases the probability of detecting the odor in a complex mixture (Arao, Suzuki, Katayama, & Yagi, 2012). As regards perceived odor intensity, colored solutions smelled orthonasally are perceived as smelling stronger than equally concentrated colorless solutions, irrespective of whether the color is appropriate for the odor (Koza, Cilmi, Dolese, & Zellner, 2005; Zellner & Kautz, 1990). This observation suggests that perceiving any color suggests the presence of an active substance, which may increase the intensity of the perceived odor. Surprisingly, colored solutions smelled retronasally do not

show any enhancement and may even show a decrease in odor intensity compared to colorless solutions (Koza et al., 2005; Zampini, Sanabria, Phillips, & Spence, 2007; Zellner & Durlach, 2003). Zellner (2013) suggests that the retronasal odor may be judged as less intense, because its perception contrasts with the orthonasal odor perceived just before the stimulus entered the mouth, and whose intensity was enhanced by the color.

Furthermore, appropriately colored odors are usually judged to be more pleasant than inappropriately colored odors (Davis, 1981; Zellner et al., 1991). This increase in hedonic appreciation has been confirmed during brain scanning (Österbauer et al., 2005). If an inappropriate color elicits an olfactory percept that deviates from the percept elicited by the odor itself, the combination of the two may smell odd. In addition, an inappropriate color may interfere with odor identification and, as a consequence, the odor will remain vague and unfamiliar (Zellner, 2013). Whether the pleasantness of congruent odor–color combinations is also higher if the odor and the color are not physically integrated remains to be determined. In the context of the present study, a color profile may be constructed for each fragrance on the basis of odor–color associations found in previous studies. This may form the basis for choosing the color of the fragrance fluid, and for choosing the color palette of the packaging and promotion materials. However, will these correspondences also have a positive effect on the liking or buying intention for the product? Although this way of working is intuitively appealing, its effectiveness has not been demonstrated empirically.

There are some indications that the latter question should be answered affirmatively. For instance, when a color fits a fragrance, their combination communicates a coherent message and is more likely to be regarded as a unitary whole. Perceived unity, in turn, is correlated with ratings of aesthetic appeal and product liking (Bell, Holbrook, & Solomon, 1991; Bone & Jantrania, 1992; Vryzer & Hutchinson, 1998). The colors of packaging are likely to generate expectations with regard to the properties of the packaging's content (Scharf & Volkmer, 2000; Shankar, Levitan, & Spence, 2010). We see this, for instance, in the words that people mention in response to packages with different colors (Ares & Deliza, 2010). And these expectations have consequences for the processing of olfactory information (Schifferstein, 2001). For example, the color of fragrance packaging may affect the perception of fragrance qualities and intensity (Gatti, Bordegoni, & Spence, 2014; Scharf & Volkmer, 2000). In addition, seeing an appropriate packaging color may help in identifying the fragrance category or gender classification (Zellner, McGarry, Mattern-McClory, & Abreu, 2008), identifying the odor's source, or naming the odor (Herz, 2003), which is likely to facilitate the further processing of the fragrance information. However, if the color creates an association with a very unpleasant odor source, the impact on the fragrance liking is likely to be negative (Herz & von Clef, 2001).

1.3. Present study

In the present study, we develop and perform an empirical test to determine whether a fragrance presented in a package with congruent colors generates higher consumer buying intentions than a fragrance in a package with incongruent (or less congruent) colors. We worked with a professional designer to develop fragrance package prototypes, and tested whether matching packages indeed yielded competitive advantage.

First, we tested whether packages composed of congruent colors are perceived as matching better with the fragrance than packages composed of incongruent colors. This test is first performed in a participant sample comparable to the sample that generated the individual odor–color associations (Section 2) and is repeated in a different sample to check for the impact of cultural differences

(Section 3). These experiments are used to select three fragrances and the corresponding packages that are most congruent. In Section 5, participants were presented with congruent and incongruent fragrance–package combinations in order to determine whether the congruent packages yielded higher buying intentions than the incongruent packages.

2. Experiment 1A

2.1. Method

Participants were presented with five different fragrances, one at a time, and were asked to evaluate how well the fragrance matched with one of five packages presented in random order.

2.1.1. Participants

Participants were recruited among students and staff of the Faculty of Industrial Design Engineering at Delft University of Technology in the Netherlands. In order to keep our sample comparable to the sample used by Schifferstein and Tanudjaja (2004), we recruited only female participants with a Dutch background, ranging in age between 18 and 40 years. The final sample consisted of 66 females, ranging in age between 18 and 36 years old (mean age 22.4 years).

2.1.2. Fragrance and color selection

For the selection of the colors and fragrances, we started out from the mean degree-of-fit ratings between fragrances and NCS color chips obtained by Schifferstein and Tanudjaja (2004). Note that these color–odor matches are restricted to the 17 colors that were included in that study. Hence, they are not necessarily the optimal color matches for these fragrances, but they are the best achievable under the constraints of the study protocol.

We selected the four fragrances that showed the largest differences in color profile (e.g., Wish, Kouros, DKNY, Miss Dior) for these 17 different colors. For each fragrance, the professional designer selected three colors that rated among the highest for this fragrance and among the lowest for the other three fragrances, and when combined together provided a harmonious color set (Table 1 and Fig. 1). Each color set was unique, implying that it contained three colors that did not appear in any of the other color sets.

In addition, we selected a unique color set for a fifth fragrance (Paris), for which the colors received high matching responses for one fragrance (Wish) and low responses for the other three fragrances. Given these five fragrances and color sets, we expected to obtain similar patterns of matching responses for the fragrances Paris and Wish, but different response patterns for each of the other fragrances Kouros, DKNY, and Miss Dior. Table 1 also describes the main colors of the original packages of the five fragrances, with their approximate NCS values (Fig. 2).

2.1.3. Package design

A professional designer created five packages that differed only with respect to the colors used. He created a standard, rectangular package (30 × 30 × 102 mm) that only contained rectangular

sections in different colors. The major color covered about 60%, the intermediate color about 25%, and the minor color about 15% of the four largest, rectangular surfaces (30 × 102 mm). In addition, the major color covered the top and bottom surfaces (30 × 30 mm). To provide a neutral, uninformative brand name, each package carried the label 1° on the minor color surface, printed in either black or white font, depending on the background color (see Fig. 1). Original NCS paper was used to create the packages, to make sure that the colors matched exactly with the NCS standard values.

2.1.4. Fragrance samples

We bought one bottle of Eau de Toilette for Kouros, DKNY, Paris, and Miss Dior original. Because Eau de Toilette was unavailable, we bought a bottle of Eau de Parfum for Wish, which is more concentrated. After purchase, fragrances were stored at about 4 °C to prevent chemical degradation.

Fragrance samples were prepared in 150 ml closable, transparent polyethylene cups. A 20 mm × 30 mm piece of white, odorless sniffing paper (Orlandi, New York) was placed in each cup. The smallest possible amount of fragrance was sprayed on the sniffing paper by pressing the nozzle of the bottle briefly. Immediately after spraying, the cup was closed with a cap to ensure that all fragrance notes (top, middle, and base) were present. Fragrance samples were prepared every day at least 1 h before the first session.

2.1.5. Procedure

The experiment was performed in a room with windows that was aired before, during and after every session day. Upon arrival, participants filled out a form with standard demographic questions and general questions describing their fragrance use and shopping behavior. Subsequently, they were presented with the first of the five fragrances. Participants were instructed to remove the plastic lid of the cup and lift the container until it was approximately 20 cm away from their nose and to use their free hand to wave the air, enabling them to smell the fragrance. Only if they were unable to smell the fragrance at a distance, they were allowed to move the container closer to their nose. Then they imagined a color palette they thought would fit to this fragrance sample. They were allowed to re-smell again after giving their nose a 20 s rest between consecutive sniffs. Subsequently, they were presented with five packages one at a time on a white background in random order. For each package they indicated how well the package matched with the fragrance using a 9-point unnumbered scale ranging from ‘very inappropriate’ on the left to ‘very appropriate’ on the right side. After evaluating all five packages for a single fragrance, they returned the fragrance sample and the completed response form, and they received the next fragrance sample, and so on. To avoid adaptation, participants waited for 60 s, they smelled their wrist or the inside of their elbow, and they took a sip of tap water between sniffing different samples.

2.1.6. Data analysis

Appropriateness ratings were analyzed by repeated-measured analyses of variance. In accordance with Stevens (2002), we

Table 1
Original colors and final selection of colors (with NCS codes) for the five fragrances with the corresponding mean color–odor degree-of-fit ratings.

	Original colors		Selected colors			Degree-of-fit ratings		
	Major	Minor	Major	Intermediate	Minor	Major	Intermediate	Minor
Miss Dior	White 0500 N	Off gray 3502 Y	Green 2565 G	Brown 6030 Y50R	Gray 4500 N	4.46	3.71	3.64
Kouros	Dark blue 8010 R70B	Medium blue 5540 R70B	Blue 3560 R80B	Light blue 1040 R80B	Black 9000 N	5.33	5.36	3.25
DKNY	White 0000 N		Yellow 0580 Y	Light yellow 0540 Y	Light green 0540 G	5.97	5.78	5.17
Wish	Blue 5540 R90B		Red 1080 R	Purple 3050 R50B	Pink 0540 R30B	5.26	4.90	6.19
Paris	Light red 0565 R		Light red 0540 R	Light orange 0540 Y40R	Light purple 1040 R50B	5.83	5.97	5.83

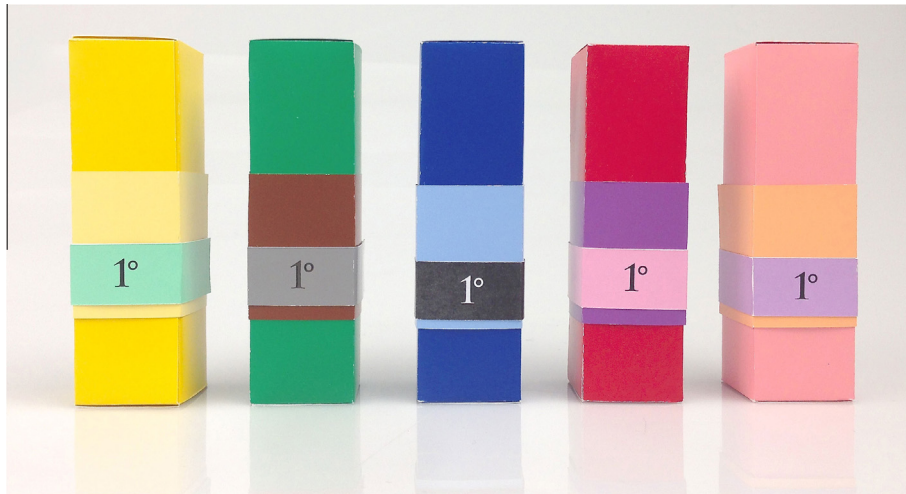


Fig. 1. Designs of the five experimental fragrance packages.



Fig. 2. Designs of the five original fragrance packages.

corrected the degrees of freedom with the Greenhouse – Geisser ε if $\varepsilon < 0.7$, and we averaged the ε values from Greenhouse – Geisser and Huynh – Feldt, when $\varepsilon > 0.7$. Differences between individual samples were investigated by a posteriori t -tests with Bonferroni adjustment.

2.2. Results

The appropriateness ratings for the 25 fragrance–package combinations were subjected to repeated measures analysis of variance (ANOVA) with Fragrance (5 levels) and Package (5 levels) as within-subjects variables. In the overall analysis of variance, we found a significant main effect of Package [$F(3, 193) = 7.1, p < 0.001$] and a Fragrance \times Package interaction [$F(9, 575) = 6.9, p < 0.001$], but no main Fragrance effect [$F(4, 260) = 1.3, p > 0.20$]. Table 2 shows the mean appropriateness ratings for the 5 fragrances and the five different packages.

To investigate these differences more closely, we performed separate ANOVAs per fragrance, with Package as within-subjects variable. With these analyses we found significant Package main effects for Kouros [$p < 0.05$], DKNY [$p < 0.05$], Wish [$p < 0.001$], and Paris [$p < 0.001$], but not for Miss Dior [$p > 0.10$]. Hence, all five packages were judged to be about equally appropriate for Miss

Table 2

Mean appropriateness ratings (\pm SE) for the five fragrances and the five packages in the Dutch sample.

	Kouros	DKNY	Wish	Paris	Miss Dior
Blue	5.24 (0.28)	5.41 (0.29)	4.39 (0.28)	4.11 (0.27)	4.76 (0.28)
Green	5.47 (0.32)	4.76 (0.30)	3.63 (0.27)	3.68 (0.26)	4.89 (0.28)
Yellow	4.50 (0.27)	5.83 (0.26)	4.77 (0.24)	5.57 (0.22)	4.47 (0.27)
Red	5.05 (0.27)	4.41 (0.25)	5.83 (0.28)	5.73 (0.27)	5.49 (0.27)
Light red	4.09 (0.30)	4.97 (0.29)	6.00 (0.30)	6.17 (0.24)	5.15 (0.32)

Dior. For Kouros and DKNY we found only one significant difference in the post hoc paired comparisons. For Kouros, the Red package was judged to be significantly more appropriate than the Light red package. For DKNY, the Yellow package was significantly more appropriate than the Red package.

For Wish and Paris, we see similar patterns for the responses to the Red, Light red, Green, and Blue packages. For these fragrances, the responses to the Red and Light red packages are highest and about the same. Responses to the Green and the Blue package are lowest, and do not differ significantly. Responses to Red and Light red packages are significantly higher than those to the Green and Blue packages [$p < 0.05$]. As regards the responses to Yellow, Wish and Paris differ. For Paris, the Yellow package is

judged to be as appropriate as the Red and Light red package, and it differs only significantly from the Green and Blue packages. For Wish, however, the Yellow package takes an intermediate position, with appropriateness ratings significantly lower than the Red and Light red packages [$p < 0.05$], significantly higher than the Green package [$p < 0.05$], but no difference with the Blue package [$p > 0.20$].

In order to evaluate the effectiveness of our manipulation, we determined whether we could predict the mean appropriateness ratings for the 25 package–odor combinations on the basis of the mean goodness-of-fit ratings for the individual color–odor combinations in a regression analysis. Because the surface areas decrease from major to minor colors, we expect the regression coefficients also to decrease in this order. In addition, the mean goodness-of-fit ratings for the Intermediate and Minor colors were generally lower than those for the Major colors, which is likely to decrease the regression coefficients (Table 1). The regression equation we found was:

$$\text{Package} = 1.69 + 0.60 * \text{Major} + 0.14 * \text{Intermediate} - 0.00 * \text{Minor} (R^2 = 0.74)$$

The equation shows that the differences are mainly explained by the differences in the major packaging colors. The regression coefficient for the major colors (0.60) is the only coefficient that differs significantly from zero [$t = 3.9$, $p < 0.01$].

3. Experiment 1B

3.1. Method

To check whether color–fragrance correspondences are culturally dependent, Experiment 1A was repeated in a sample of US university students. Materials and procedure were similar to those described in Section 2. Any important differences are mentioned below.

3.1.1. Participants

Participants were recruited among students and staff on the campus of Brigham Young University in Utah, USA. We recruited only female participants ranging in age between 18 and 40 years. The final sample consisted of 80 females, ranging in age between 18 and 30 years old (mean age 20.8 years). The vast majority of the sample (94%) grew up in the USA.

3.1.2. Procedure

The experiment was performed outside, on a table in a quiet and shady part of a public area on the BYU campus, near to the location where participants were recruited. Instructions were identical to Section 2.

3.2. Results

The appropriateness ratings for the 25 fragrance–package combinations were subjected to repeated measures analysis of variance (ANOVA) with Fragrance (5 levels) and Package (5 levels) as within-subjects variables. In the overall analysis of variance, we found a significant main effect of Fragrance [$F(4, 316) = 3.6$, $p < 0.01$] and a Fragrance \times Package interaction [$F(10, 766) = 7.6$, $p < 0.001$], but no main Package effect [$F(3, 258) = 1.1$, $p > 0.20$]. Table 3 shows the mean appropriateness ratings for the five fragrances and the five different packages.

To investigate these differences more closely, we performed separate ANOVAs per fragrance, with Package as within-subjects variable. With these analyses we found significant Package main effects for Kouros, Wish, and Paris [all $p < 0.001$], but not for Miss Dior [$p > 0.05$] and DKNY [$p > 0.20$]. For Kouros, the Green and Blue packages fit significantly better than the Red, Light red, and Yellow packages. In contrast, for Wish, the Red, Light red and

Table 3

Mean appropriateness ratings (\pm SE) for the five fragrances and the five packages in the US sample.

	Kouros	DKNY	Wish	Paris	Miss Dior
Blue	6.09 (0.23)	4.98 (0.25)	4.75 (0.25)	4.68 (0.25)	5.43 (0.26)
Green	5.84 (0.27)	5.01 (0.29)	4.23 (0.27)	4.51 (0.27)	5.40 (0.27)
Yellow	4.36 (0.27)	5.35 (0.26)	5.60 (0.26)	5.15 (0.27)	4.54 (0.25)
Red	4.63 (0.27)	5.05 (0.26)	5.89 (0.25)	5.89 (0.24)	4.85 (0.25)
Light red	3.75 (0.28)	5.14 (0.29)	6.25 (0.29)	5.94 (0.24)	4.70 (0.29)

Yellow packages fit better than the Green one. In addition, for Wish the Light red package fits significantly better than the Blue package. Similar to the previous study, results are comparable for Wish and Paris: for Paris, the Red and the Light red packages fit significantly better than the Blue and Green packages [all $p < 0.05$].

We also tested whether we could predict the mean appropriateness ratings of the US respondents for the 25 package–odor combinations on the basis of the mean goodness-of-fit ratings of a group of Dutch respondents for the individual color–odor combinations. The regression equation we found was:

$$\text{Package} = 2.28 + 0.68 * \text{Major} + 0.06 * \text{Intermediate} - 0.13 * \text{Minor} (R^2 = 0.59)$$

Again, the equation shows that the differences are mainly explained by the differences in the major packaging colors: only the coefficient for the Major color (0.68) differs significantly from zero [$t = 3.9$, $p < 0.01$].

4. Discussion

Sections 2 and 3 show large overlap in outcomes. However, there are also some differences that can be noted. The largest deviations in mean responses between the two participant samples are found for Kouros in the Blue package, Wish in the Yellow package, and Paris in the Green one. It is interesting to see that the Dutch women clearly distinguish between DKNY and Wish, which they associate with the colors Yellow and Red, respectively, whereas the US women provide very similar responses for these two fragrances. It is unclear where the differences between samples stem from. As stated in Section 1, some color–odor correspondences may have a semantic or a hedonic origin. Both these mechanisms may depend on the exposure people have had to specific stimuli or stimulus combinations in different cultures. For instance, being exposed to different odor–color combinations may explain why people have different flavor expectations with different colors (Shankar et al., 2010). In addition, people tend to appreciate stimuli that they have encountered before, the so-called mere exposure effect (Bornstein, 1989; Zajonc, 1968). Hence, familiar items also tend to be liked more than unfamiliar ones.

In both studies the matching responses for Miss Dior were not significantly higher for the Green package than for the other packages. Perhaps this is not surprising, given that the average degree-of-fit ratings on which they were based (Table 1) were also relatively low, compared to the ratings for the other packages. Nonetheless, Miss Dior's ratings for these three colors were among the highest and were rather typical in comparison to the group of 14 fragrance samples used in the original study (Schifferstein & Tanudjaja, 2004). The current outcomes suggest, however, that it is not the relative position in the group, but rather the absolute degree-of-fit rating of a color that will determine its appropriateness for designing a fragrance package.

Our regression analyses suggest that the appropriateness of a fragrance package is affected mainly by the major color used for the design of the packaging, whereas the additional colors have smaller or no effects. The minor effects of the additional colors

may not only have been due to the smaller sizes of the respective packaging areas, but also to the generally lower goodness-of-fit of these colors. Our outcomes raise the question whether the appropriateness of a fragrance package may be exclusively dominated by the main color, while the appropriateness of the additional colors is more or less neglected. For this discussion, it is important to take the perceived relationships between the color elements into account. Our designer deliberately created packages in which three colors formed harmonious combinations. By using only harmonious combinations, the intermediate and minor colors acted as supporting colors, and were unable to stand out from the package and grab attention. Hence, the dominance of the major color may have been partly due to the way in which we selected color combinations for the package designs.

One variable that may have interfered with our study concerns the original packaging of the fragrances: participants who were familiar with a fragrance and its package may have recognized the fragrance during the experiment, and the associations with the colors of the original packages may have affected their judgments. However, we think it is unlikely that such associations have had a substantial impact on our findings. First of all, we selected fragrances for the study that were available on the market, but that were not very popular, hence reducing the chance of recognition. Furthermore, only very few participants indicated that they recognized a fragrance during the experiment, and if we checked whether they correctly identified the fragrance, they generally reported a fragrance name that was not part of the study. In addition, our data do not show any signs of color associations with the original packages. If fragrance recognition had affected our responses, we would expect fragrances with similar original packages to elicit similar color associations. For instance, Wish and Kouros both are sold in Dark blue packages, but whereas the odor of Kouros is matched to Blue and Green packages, the odor of Wish is mainly matched with Red and Light-red packages. Analogously, Miss Dior and DKNY are both sold in White packages, but the odor of Miss Dior obtains the highest matching ratings for the Red package, whereas the odor of DKNY is mostly associated with a Yellow package. Therefore, we do not see any evidence that associations with the original package have affected the current results.

In terms of overlap, both studies show that the degree-of-fit for the package can be predicted well by the package's major color, for which the degree-of-fit was determined in a previous study (Schifferstein & Tanudjaja, 2004). The R^2 value of this regression analysis in Section 3 is a bit lower than in Section 2, which may be due to cultural differences. In addition, the tendencies in mean color–fragrance appropriateness responses are quite similar in the two studies. Because the overlap between the studies is large and we find reliable differences in package–odor congruency ratings in the US sample, we are confident that a test in the US will enable us to determine whether congruent combinations are liked better, leading to higher purchase intentions among consumers.

5. Experiment 2

5.1. Method

In order to test whether people prefer to buy fragrance–package combinations that match well, we need to have fragrances that match well with one package, but do not match with the other packages. Therefore, we selected three fragrances and three packages with distinctive matching profiles out of the five fragrances and five packages investigated in Sections 2 and 3, together forming nine possible fragrance–package combinations. Each participant in Section 5 was presented with three of these combinations. Participants evaluated the extent to which they appreciated these combinations by reporting their purchase intentions. We expect that congruent combinations will be appreciated more than incongruent combinations.

5.1.1. Participants

Participants were recruited among students and staff of Brigham Young University in Utah, USA using the same selection criteria as in Section 3. The sample consisted of 109 females, ranging in age between 18 and 25 years old (mean age 19.7).

5.1.2. Stimulus selection

Wish and Paris obtained high matching ratings for Red and Light red packages, and low ratings for Blue and Green packages. The Yellow package was also rated quite high, although the Dutch sample found that it matched Wish somewhat less, resulting in an intermediate position. For Kouros the Dutch sample only exhibited a difference between the Red and Light red package. The US sample, however, exhibited consistently higher matching scores for the Blue and the Green packages than for the Red, Light red and Yellow packages. Although package differences for DKNY were not significant in the US sample, in the Dutch sample the mean matching response for the Red package was significantly lower than for Yellow.

The overview above suggests that the Red and Light red packages exhibit quite similar patterns, just like the Green and Blue packages, whereas the responses to Yellow seem to be more unique. Analogously, the fragrances Wish and Paris exhibit similar patterns, whereas Kouros and DKNY show different patterns. Miss Dior is not an interesting candidate for Section 5, because it did not yield any significantly different matching scores between packages. Because we are looking for fragrance–package combinations that evoke different matching responses, stimuli that provide similar response patterns are not so interesting for inclusion in the next experiment. On the basis of these considerations, we decided to select the Red, Green and Yellow packages, and Wish, DKNY and Kouros as fragrances. The three congruent combinations are Wish – Red, DKNY – Yellow, and Kouros – Green, whereas all other combinations are considered incongruent.

5.1.3. Experimental design

Each participant received three fragrance–package combinations in random order. Each fragrance and each package were only presented once. There are six possible ways in which the three fragrances can be combined with the three packages. An approximately equal number of participants ($n=18$ or 19) were presented with each of these sets, resulting in a total sample size N of 109 participants.

5.1.4. Procedure

Experiment 2 was performed in a quiet area of a very large open space at ambient temperature, with sufficient lighting and air movement, where a low number of students passed. After a short introduction, participants were presented with the first fragrance–package combination and they smelled the sample following the instructions described under Section 2. Then they filled out a questionnaire with six questions. The participants indicated the likelihood that they would purchase this fragrance in this package (1) for themselves, (2) for a male friend or relative, and (3) for a female friend or relative on a 9-point unnumbered scale ranging from 'very unlikely' to 'very likely'. They also indicated how well the package and the colors matched with the fragrance using a 9-point scale ranging from 'very inappropriate' to 'very appropriate'. In addition, they indicated how well they liked the scent and the package on a 9-point scale with left anchor 'do not like at all' and right anchor 'like very well'. After evaluating the first fragrance–package combination, they returned the combination and the completed response form, and received the next combination. After the experiment, participants filled out a form with standard demographic questions and general questions describing their fragrance use and shopping behavior.

5.1.5. Data analysis

Data were analyzed by two-way ANOVA with Fragrance (with levels Wish, Kouros, DKNY) and Package (with levels Red, Green, Yellow) as fixed factors. In the data analysis we disregarded that each subject had evaluated three combinations during a session. Instead, we analyzed the data as if each person had evaluated one fragrance–package combination only.

5.2. Results

Ideally, if respondents can separately assess their evaluations of fragrance and packaging independently, we expect to find main effects of Fragrance and Packaging for liking evaluations of the fragrances and the packages, respectively, without any other effects. This is indeed what we found in two-way ANOVAs for these dependent variables. We found a Package main effect for the responses on ‘How well do you like this package?’ [$F(2, 318) = 11.6, p < 0.001$], but no other effects [$p > 0.20$]. In addition, we found a Fragrance main effect for ‘How well do you like this scent?’ [$F(2, 318) = 23.9, p < 0.001$] and no other effects [$p > 0.20$].

In contrast, we expect to find a Fragrance \times Package interaction for ‘How well do the package and the colors match the fragrance?’ because this question directly addresses the interrelationships between fragrance and packaging. Indeed, we found a significant two-way interaction [$F(4, 318) = 6.7, p < 0.001$]. In addition, this dependent variable showed a significant Fragrance main effect [$F(2, 318) = 4.6, p < 0.05$], but no Package main effect [$p > 0.10$]. The mean matching ratings roughly concur with our predictions that the Red package matches best with Wish, the Yellow with DKNY, and the Green package with Kouros (see Table 4).

Most interestingly, if the degree of congruence between fragrances and packages affects the participants’ buying intentions, we expect to find a significant Fragrance \times Package interaction for the various types of buying intentions as well. However, none of the additional dependent variables show a two-way interaction [all $F(4, 318) < 1.0$ and all $p > 0.20$]. For the likelihood that the participant would buy the fragrance for herself, we did find a significant main effect for Fragrance [$F(2, 318) = 28.7, p < 0.001$], while the Package main effect just failed to reach significance [$F(2, 318) = 2.7, p = 0.069$]. The likelihood that she would buy the fragrance for someone else showed main effects for Fragrance and Package, both for male [$F(2, 318) = 33.0, p < 0.001$ for Fragrance and $F(2, 318) = 11.4, p < 0.001$ for Package] and female [$F(2, 318) = 11.6, p < 0.001$ for Fragrance and $F(2, 318) = 16.4, p < 0.001$ for Package] friends or relatives.

Table 4

Mean matching ratings (\pm SE) between the three fragrances and the three packages in Section 5 (9-point scale).

	Kouros	DKNY	Wish
Green	6.57 (0.37)	6.19 (0.38)	4.33 (0.38)
Yellow	5.56 (0.38)	7.11 (0.37)	6.36 (0.38)
Red	5.08 (0.38)	6.25 (0.38)	6.38 (0.37)

Table 5

Mean ratings (\pm SE) for liking and purchase intentions for the fragrances and the packages (9 point scale).

	Liking	Purchase for oneself	Purchase for male	Purchase for female
<i>Fragrance</i>				
DKNY	5.90 (0.23)	4.83 (0.22)	5.19 (0.23)	3.19 (0.22)
Kouros	4.24 (0.23)	3.08 (0.22)	3.45 (0.23)	4.12 (0.22)
Wish	6.34 (0.23)	5.34 (0.22)	5.98 (0.23)	2.63 (0.22)
<i>Package</i>				
Red	6.17 (0.20)	4.49 (0.22)	5.32 (0.23)	2.73 (0.22)
Yellow	6.18 (0.20)	4.74 (0.22)	5.31 (0.23)	2.86 (0.22)
Green	5.01 (0.20)	4.02 (0.22)	3.99 (0.23)	4.34 (0.22)

These analyses suggest that participants are able to evaluate whether they like a fragrance, independently from the package with which it is presented. Similarly, participants are able to evaluate whether they like a package independently from the fragrance with which it is presented. In addition, they can judge which fragrances match with which packages in a way that is roughly consistent with results from previous experiments. However, the degree of matching does not seem to influence their degree of liking for a particular fragrance or a particular package, nor does it seem to affect their buying intentions for the various fragrance–package combinations.

For all fragrance–package combinations, participants indicated the likelihood that they would purchase this fragrance in this package for themselves, for a male friend, and for a female friend. From these data, we calculated the mean buying intentions for the fragrance (averaged over packages) and the package (averaged over fragrances) (Table 5). In addition, participants indicated how much they liked the fragrance and the package. These means are also reported in Table 5. These means indicate that the participants tend to like the fragrances Wish and DKNY, and the packages Red and Yellow. In addition, they might consider to buy Wish for a male friend, but otherwise their buying intentions are rather low, especially if it concerns buying a fragrance for a female friend.

In order to determine what affects the buying intentions for various fragrance–package combinations, we performed stepwise regression analyses, in which we used buying intention as the dependent variable and the liking rating for the fragrance, the liking rating for the package, and the degree of match between fragrance and package as the three explanatory variables. These analyses show that the intention to buy a fragrance for themselves was predicted by the liking for the fragrance ($\beta = 0.68, p < 0.001$) and the package ($\beta = 0.12, p < 0.01$). Analogously, when buying a fragrance for a male, their intention was predicted by her liking for the fragrance ($\beta = 0.61, p < 0.001$) and the package ($\beta = 0.16, p < 0.01$). However, when buying a fragrance for a female, their intention was much less related to her liking scores. Only the coefficient for the fragrance ($\beta = 0.14, p < 0.05$) contributed significantly. Given that fragrance and package had significant impact on buying intentions for another female in the ANOVA, the intentions of buying a fragrance for a female seem to be driven by factors other than the liking for the individual fragrances and packages. The degree of fragrance–package match did not pass the criterion for entering the equation in any of these analyses ($p > 0.05$).

6. General discussion

6.1. Is preference for a complex stimulus determined by the liking of the individual (independent) components, or by the degree of congruence between the different elements?

In the present study, we found that buying intentions for the fragrance–package combinations were mainly related to the liking scores for the fragrances and, to a lesser extent, to the liking scores

for the packages. The degree of matching did not seem to play a role in determining the buying intentions.

Finding no effect of matching is not uncommon in studies involving smells. For instance, Schifferstein and Michaut (2002) investigated the effect of adding congruent or incongruent pleasant odors on product evaluations for food, personal care and household products. These authors also found no effect of the degree of odor-product congruency on overall product evaluations and buying intentions. They speculated that odor-product congruency might be more important during product trial than during product selection. Ludden and Schifferstein (2009) found no effect of the degree of (in)appropriateness of odors on the affective evaluations of industrial products that normally do not carry an odor. Analogously, Porcherot, Delplanque, Gaudreau, and Cayeux (2013) concluded that the optimization of color–perfume pairing did not result in a significant increase of the subjective affective responses to fabric softeners. Similar to the present study, their study underlines the predominant role of perfume in the emotional response and a more limited influence of color.

Our findings seem to contradict those obtained by Bone and Jantrania (1992), who found higher pleasantness ratings for appropriately scented household cleansers and sunscreen lotions than for inappropriately scented and unscented ones. One difference with their study lies in the type of product: whereas the sole primary function of a fragrance is to produce a nice smell, a cleaning or personal care product is also expected to clean properly or to nourish the user's skin, respectively. Hence, the focus on the pleasantness of the fragrance is likely to be more dominant for the fine fragrances than for the other products. Another difference is that we instructed participants to judge whether the fragrances had a nice smell. By directing their attention explicitly to the properties of the smell, we may have increased the salience and the number of thoughts directed to the fragrance, while in Bone and Jantrania's study participants may have reacted more holistically to the product offering. When people evaluate complex products, they may employ various strategies that differ in the extent to which they pay attention to and actively process the various product aspects (Carbon & Leder, 2005; Sujan, 1985). Different information processing strategies usually result in different product evaluations. Furthermore, please note that some studies have shown that the liking for individual product aspects is not always a good predictor of the preference for a complex product (Schifferstein, Otten, Thoolen, & Hekkert, 2010).

6.2. Considerations in fragrance buying for women and men

We were quite surprised by the finding that women who buy a fragrance for a female friend or relative do not seem to choose a fragrance or package they would like. In general, the intentions of buying a fragrance for another female were extremely low. In addition, all regression coefficients were below 0.15, which is rather low compared to the values found when buying fragrances for themselves or for male friends. To clarify these findings, we interviewed 12 women from the US target group about their motivations for buying specific fragrances. These interviews indicated that it is uncommon for women to buy fragrances for other women. Women see fragrances as important parts of their personal identity. They consider them to be very personal and intimate. They may use the same fragrance for many years, and some women keep their fragrance choice a secret, so that their friends will not wear the same fragrance. Therefore, they would not purchase the same fragrance for a female friend.

In addition, giving someone a fragrance might suggest that they need one because they smell bad. This could cause unwanted questions to occur between friends. Women prefer to avoid that possibility and look for a gift that is safer to share. In cases where

women do give their friends fragrances, it may concern fragrances that are no longer working for themselves and that are passed on to their friends, in hopes that they will work for them. Alternatively, women may appreciate the personality or preference differences between themselves and their friends, and may decide to give them a fragrance that has a different character than what they would buy for themselves. Fragrance categories (e.g., fruity, spicy, citrus, floral) may play an important role in this process: Some women identified themselves as belonging to a certain category and they would consider purchasing only fragrances from this category for themselves. Consequently, they would not recommend these fragrances for their female best friends.

In contrast, women commonly buy fragrances for men, because they want to influence what fragrances men wear. Assuming the man is a boyfriend or spouse, they want him to wear something they also like, because they will be together often. The gender classification of the fragrance can be an important determinant of the degree of liking: While women may like a certain fragrance for themselves, they may like other fragrances for men. In the line of the present study, it is interesting to note that Zellner et al. (2008) showed that thinking of the masculinity/femininity of a fragrance influenced the selection of colors that correspond to these odors. In the present study, we did not inform participants whether fragrances were created for women or men. Nonetheless, some participants may have guessed that some fragrances (e.g., Kouros) were male fragrances and may have adapted their color choices. It will be interesting to take gender classification into account in future studies on fragrance packaging design.

6.3. Possible implications for design

Our findings suggest that color–fragrance associations currently do not influence consumer preferences for fragrances. Probably, when fragrance companies make design decisions about package, fluid, and bottle colors, they more often base their choices on trends or traditions in style, rather than on empirical evidence on intuitive associations with other product characteristics. Typically, designers may make several packaging proposals based mainly on artistic expression, differentiation from competing brands and continuation of the own brand style, and let an audience pick the one that is most appropriate for introduction on the market.

However, even though we found no effects of color matching on product buying intentions in the present study, we think that it would be unwise and premature to conclude that color appropriateness is unimportant for fragrance packaging design. Odors do not only show correspondences with colors, but also with shape (Becker, van Rompay, Schifferstein, & Galetzka, 2011; Hanson-Vaux, Crisinel, & Spence, 2013) and abstract symbols (Seo et al., 2010). This may provide additional direction for design strategies. Currently, brand, packaging, and bottle shape are the first attractors when entering a fragrance store. Therefore, congruency among the stimuli in their design is important to construct a coherent fragrance message and to raise the correct expectations. The packaging design is likely to attract consumers who are looking for fragrances with specific characteristics, and if these expectations are not met, consumers are less likely to find the fragrance they look for.

Alternatively, we may start to develop an alternative way in which fragrances are created and sold. For instance, to follow some women's desire to hide the identity of their fragrance, we might develop stores where fragrances are available in bulk containers, aligned on a wall by category. Consumers may select an unbranded bottle from a variety of shapes to then fill with whatever scent they want (or mix their own). In this case, the bottle would no longer

identify their fragrance, and they could also explore mixing their own bespoke fragrances.

Acknowledgment

The authors would like to thank J. Drew Smith for his contribution to data collection and data processing of Experiments 1B and 2.

References

- Arao, M., Suzuki, M., Katayama, J., & Yagi, A. (2012). An odorant congruent with a colour cue is selectively perceived in an odour mixture. *Perception*, 41(4), 474–482.
- Ares, G., & Deliza, R. (2010). Studying the influence of package shape and colour on consumer expectations of milk desserts using word association and conjoint analysis. *Food Quality and Preference*, 21(8), 930–937. <http://dx.doi.org/10.1016/j.foodqual.2010.03.006>.
- Becker, L., van Rompay, T. J. L., Schifferstein, H. N. J., & Galetzka, M. (2011). Tough package, strong taste: The influence of packaging design on taste impressions and product evaluations. *Food Quality and Preference*, 22, 17–23.
- Bell, S. S., Holbrook, M. B., & Solomon, M. R. (1991). Combining esthetic and social value to explain preferences for product styles with the incorporation of personality and ensemble effects. *Journal of Social Behavior and Personality*, 6, 243–274.
- Bone, P. F., & Jantrania, S. (1992). Olfaction as a cue for product quality. *Marketing Letters*, 3, 289–296.
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research 1968–1987. *Psychological Bulletin*, 106, 265–289.
- Carbon, C. C., & Leder, H. (2005). The repeated evaluation technique (RET). A method to capture dynamic effects of innovativeness and attractiveness. *Applied Cognitive Psychology*, 19, 587–601.
- Davis, R. G. (1981). The role of nonolfactory context cues in odor identification. *Perception & Psychophysics*, 30, 83–89.
- Demattè, M. L., Sanabria, D., & Spence, C. (2006). Cross-modal associations between odors and colors. *Chemical Senses*, 31, 531–538.
- Engen, T. (1982). *The perception of odors*. New York: Academic Press.
- Gatti, E., Bordegoni, M., & Spence, C. (2014). Investigating the influence of colour, weight, and fragrance intensity on the perception of liquid bath soap: An experimental study. *Food Quality and Preference*, 31, 56–64.
- Gilbert, A. N., Martin, R., & Kemp, S. E. (1996). Cross-modal correspondence between vision and olfaction: The color of smells. *American Journal of Psychology*, 109, 335–351.
- Hanson-Vaux, G., Crisinel, A.-S., & Spence, C. (2013). Smelling shapes: Crossmodal correspondences between odors and shapes. *Chemical Senses*, 38(2), 161–166.
- Herz, R. S. (2003). The effect of verbal context on olfactory perception. *Journal of Experimental Psychology: General*, 132, 595–606.
- Herz, R. S., & von Clef, J. (2001). The influence of verbal labeling on the perception of odors: Evidence for olfactory illusions? *Perception*, 30, 381–391.
- Kemp, S. E., & Gilbert, A. N. (1997). Odor intensity and color lightness are correlated dimensions. *American Journal of Psychology*, 110, 35–46.
- Kim, Y. J. (2013). Can eyes smell? Cross-modal correspondences between color hue-tone and fragrance family. *Color Research & Application*, 38(2), 139–156.
- Koza, B. J., Cilmi, A., Dolese, M., & Zellner, D. A. (2005). Color enhances orthonasal olfactory intensity and reduces retronasal olfactory intensity. *Chemical Senses*, 30(8), 643–649.
- Ludden, G. D. S., & Schifferstein, H. N. J. (2009). Should Mary smell like biscuit? Investigating scent in product design. *International Journal of Design*, 3(3), 1–12.
- Maric, Y., & Jacquot, M. (2013). Contribution to understanding odour-colour associations. *Food Quality and Preference*, 27(2), 191–195.
- Österbauer, R. A., Matthews, P. M., Jenkinson, M., Beckmann, C. F., Hansen, P. C., & Calvert, G. A. (2005). Color of scents: Chromatic stimuli modulate odor responses in the human brain. *Journal of Neurophysiology*, 93, 3434–3441.
- Porcherot, C., Delplanque, S., Gaudreau, N., & Cayeux, I. (2013). Seeing, smelling, feeling! Is there an influence of color on subjective affective responses to perfumed fabric softeners? *Food Quality and Preference*, 27(2), 161–169.
- Scharf, A., & Volkmer, H. P. (2000). The impact of olfactory product expectations on the olfactory product experience. *Food Quality and Preference*, 11, 497–503.
- Schifferstein, H. N. J. (2001). Effects of product beliefs on product perception and liking. In L. Frewer, E. Risvik, & H. Schifferstein (Eds.), *Food, people and society: A European perspective of consumers' food choices* (pp. 73–96). Berlin: Springer Verlag.
- Schifferstein, H. N. J., & Michaut, A. M. K. (2002). Effects of appropriate and inappropriate odors on product evaluations. *Perceptual and Motor Skills*, 95, 1199–1214.
- Schifferstein, H. N. J., Otten, J. J., Thoolen, F., & Hekkert, P. (2010). An experimental approach to assess sensory dominance in a product development context. *Journal of Design Research*, 8(2), 119–144.
- Schifferstein, H. N. J., & Tanudjaja, I. (2004). Visualising fragrances through colours: The mediating role of emotions. *Perception*, 33, 1249–1266.
- Seo, H.-S., Arshamian, A., Schemmer, K., Scheer, I., Sander, T., Ritter, G., et al. (2010). Cross-modal integration between odors and abstract symbols. *Neuroscience Letters*, 478(3), 175–178.
- Shankar, M. U., Levitan, C. A., & Spence, C. (2010). Grape expectations: The role of cognitive influences in color-flavor interactions. *Consciousness and Cognition*, 19(1), 380–390.
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.
- Stevenson, R. J., Rich, A., & Russell, A. (2012). The nature and origin of cross-modal associations to odours. *Perception*, 41(5), 606–619.
- Sujan, M. (1985). Consumer knowledge: Effects on evaluation strategies mediating consumer judgments. *Journal of Consumer Research*, 12, 31–46.
- Verzyer, R. W., & Hutchinson, J. W. (1998). The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of Consumer Research*, 24, 374–394.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, Monograph Supplement*, 9, 1–27.
- Zampini, M., Sanabria, D., Phillips, N., & Spence, C. (2007). The multisensory perception of flavor: Assessing the influence of color cues on flavor discrimination responses. *Food Quality and Preference*, 18(7), 975–984.
- Zellner, D. A. (2013). Color-odor interactions: A review and model. *Chemosensory Perception*, 6(4), 155–169.
- Zellner, D. A., Bartoli, A. M., & Eckard, R. (1991). Influence of color on odor identification and liking ratings. *American Journal of Psychology*, 104, 547–561.
- Zellner, D. A., & Durlach, P. (2003). Effect of color on expected and experienced refreshment, intensity, and liking of beverages. *American Journal of Psychology*, 116, 633–647.
- Zellner, D. A., & Kautz, M. A. (1990). Color affects perceived odor intensity. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 391–397.
- Zellner, D. A., McGarry, A., Mattern-McClory, R., & Abreu, D. (2008). Masculinity/femininity of fine fragrances affects color-odor correspondences: A case for cognitions influencing cross-modal correspondences. *Chemical Senses*, 33(2), 211–222.