Meat Science 188 (2022) 108766

Contents lists available at ScienceDirect

Meat Science

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

Human perception of color differences using computer vision system measurements of raw pork loin

Brianne A. Altmann [a,*], Jan Gertheiss [b], Igor Tomasevic [c], Christina Engelkes [a], Thibaud Glaesener [a], Jule Meyer [a], Alina Schäfer [a], Richard Wiesen [a], Daniel Mörlein [a]

[a] Department of Animal Sciences, Faculty of Agricultural Sciences, University of Goettingen, Kellnerweg 6, 37077 Goettingen, Germany
[b] Department of Mathematics and Statistics, Helmut Schmidt University, Holstenhofweg 85, 22043, Hamburg, Germany
[c] Food Technology and Biochemistry Department, Faculty of Agriculture, University of Belgrade, Nemanjina 6, 11080 Belgrade, Serbia

ARTICLE INFO

Keywords:
Discrimination testing
Triange test
Color preference
Food appearance
Meat color

ABSTRACT

In the food industry, product color plays an important role in influencing consumer choices. Yet, there remains little research on the human ability to perceive differences in product color; therefore, preference testing is subjective rather than based on quantitative colors. Using a de-centralized computer-aided systematic discrimination testing method, we ascertain consumers' ability to discern between systematically varied colors. As a case study, the colors represent the color variability of fresh pork as measured by a computer vision system. Our results indicate that a total color difference (ΔE) of approximately 1 is discriminable by consumers. Furthermore, we ascertain that a change in color along the b*-axis (yellowness) in CIELAB color space is most discernable, followed by the a*-axis (redness) and then the L*-axis (lightness). As developed, our web-based discrimination testing approach allows for large scale evaluation of human color perception, while these quantitative findings on meat color discrimination are of value for future research on consumer preferences of meat color and beyond.

1. Introduction

Product color is of great importance for producers and consumers in determining food quality. For example, lean muscle (meat) color is employed as an indicator of meat quality. Primarily, color is monitored as an indicator for pale, soft, exudative (PSE) or dark, firm, dry (DFD) meat (Adzitey & Nurul, 2011), as color measurements are non-destructive and quick to perform. This ensures that such products with inferior quality characteristics can be identified prior to retail. There are numerous instrumental methods applied in the meat industry to monitor color; most commonly objective measurements are taken using a spectrophotometer or one of many computer vision systems and values are reported using the 3D CIELAB color space (Mancini & Hunt, 2005; Tapp, Yancey, & Apple, 2011). In meat science, instrumental meat color is most often monitored in fresh pork (Tapp et al., 2011), likely due to its status as neither "white" nor "red meat" and the long-standing problem of PSE.

Although color may be monitored prior to retail, as a natural product, the color of meat still varies widely in the marketplace (Mörlein, Link, Werner, & Wicke, 2007). Subsequently, consumers respond to these variations in color (Ngapo, Martin, & Dransfield, 2007), because product color remains one of the first and decisive sensory characteristics evaluated by consumers during retail (Ngapo, Rubio Lozano, & Braña Varela, 2018; Tomasevic, Djekic, Font-i-Furnols, Terjung, & Lorenzo, 2021). Based on the color, consumers immediately make assumptions about other product characteristics such as freshness, husbandry conditions, processing steps and nutritional value (Kennedy, Stewart-Knox, Mitchell, & Thurnham, 2004; Ngapo et al., 2007). However, there remains limited information concerning the average consumer's ability to perceive measurable objective (instrumental) differences in color, in general, let alone along the spectrum of meat colors. Most studies concerning consumer perception of meat color focus on determining preferences for subjective colors (e.g., pink vs. red), making the comparisons between studies difficult and restricting their application for industry.

For example, Ngapo et al. (2007) have explored preferences for pork search attributes, one of which was pork color, in an international survey covering over 11,000 consumers across 22 countries. They used

* Corresponding author at: Kellnerweg 6, 37077 Göttingen, Germany.
E-mail addresses: brianne.altmann@agr.uni-goettingen.de (B.A. Altmann), jan.gertheiss@hsu-hh.de (J. Gertheiss), tbigor@agrif.bg.ac.rs (I. Tomasevic), daniel.moerlein@uni-goettingen.de (D. Mörlein).

https://doi.org/10.1016/j.meatsci.2022.108766
Received 9 November 2021; Received in revised form 4 February 2022; Accepted 6 February 2022
Available online 9 February 2022

computer-modified pork chop photos with two distinct colors ("dark red" and "light red") to show that color preferences do exist worldwide based on demographic factors, such as consumption frequency, household-makeup, age, and educational level. More recently, Lusk et al. (2018) investigated US consumers' preferences for pork color, based on quality grade labels being introduce by industry, using product photos with varying (3 levels of) meat colors. The authors conclude that consumers' preferences for meat color are heterogeneous; and despite a strong preference for redder pork chops, there exists a niche of consumers who prefer paler pork chops in the USA. The majority prefer redder pork. Based on these consumer preferences, a common goal for industry is to retail meat with an intense red hue.

Based on such subjective consumer studies and assumptions, the meat sector has invested in production and grading schemes (Lusk et al., 2018; Ngapo, Riendeau, Laberge, & Fortin, 2012) as well as technological methods (Mancini & Hunt, 2005) to improve meat color for consumers. For example, meat packaging technology, such as highly oxygenated modified atmosphere packaging, is often employed to increase red hues in raw meat products (McMillin, 2008). Yet the question remains: what constitutes "redder" pork chops, i.e., at what numerical difference in objective color measurements is a consumer able to discriminate two products? Or, more general, to what extend must two color shades vary that the human eye perceives them being different?

As briefly explained, colors are most often measured in CIELAB color space and by means of an equation derived from the three color axes, L* (lightness), a* (redness) and b* (yellowness), the so-called "total color difference" ΔE-value can be calculated to express the difference between two colors (Chen, Wardman, & Smith, 2002). There exist numerous studies on the visual perceptibility of color differences based on presentation method (Liu et al., 2018; Pointer, Attridge, & Jacobson, 2002a, 2002b), and dentistry often employs the concept of ΔE to assess tooth health and whiteness (La Rosa et al., 2019; Thoma et al., 2016). These studies often concentrate on the overall difference (ΔE) and visual perception between multiple shades of white or how items are presented. However, no studies exist concerning meat products and the influence of color direction. Understanding to what extent changes along the L*, a*, or b* axes individually influence the discrimination of color is crucial for the food industry to be able to efficiently meet consumer preferences in the future.

This study is based on the research of Tomasevic et al. (2019), who explain the advantages of their computer vision system (CVS) against spectrophotometer measurements in meat products. The authors prove that the color values determined with a spectrophotometer are far inferior to the CVS. That is, when used to reproduce colors, CVS captured values reproduce the initial product color much more closely (Tomasevic et al., 2019). Using this improved technology to capture and reproduce pork product colors, we assessed the color variability of pork and then created a systematically varied space of colors to be used for systematic discrimination testing. To study the consumers' ability to distinguish between colors, we then innovatively employed large scale computer-aided discrimination testing based on a design of experiment (DoE).

The current study tests two hypotheses:

1) based on the conclusions of Stokes et al. (1992), pictures where colors can be reproduced with a ΔE less than 2.2 are not discernably different. We assume that with a more monochrome item, such as fresh meat this threshold will still apply for consumers trying to discern differences in fresh pork color;

2) in accordance with meat science findings that a* and b* values correlate with visual perception of meat redness (Zhu & Brewer, 1999), we hypothesize that small changes along either the a* (Δa) or b* (Δb) axes effect a consumers' ability to discern between two fresh pork colors; whereas a larger difference along the L* axis (ΔL) is necessary in order for colors to be visually distinguishable.

2. Materials & methods

This study obtained ethical approval from the University of Goettingen Ethics Committee (Nr. 6/02.20-Altmann) prior to data collection. In addition, informed consent was obtained from all participants; participants were also notified that they could exit the experiment at any time. All data were collected anonymously.

2.1. Pork color determination using a computer vision system

Average pork color was determined as mean L*, a* and b* values measured across 15 pork *Longissimus thoracis et lumborum* muscles obtained from 15 different Serbian pork meat producers. The samples were chosen to obtain a variability in terms of composition, structure and color. The age of the animal was approximately 6 months and the samples were analyzed about 4 days post mortem. All samples were checked for pH using a pH meter (Testo 205, Testo, USA) equipped with a pH probe and thermometer penetration tip (Testo 0650 2051, Testo USA) in order to exclude PSE or DFD cases. The pH meter was calibrated at three points using DAkkS-certified (Deutsche Akkreditierungsstelle) calibration solutions and according to the manufacturer calibration procedure. The instrument was equipped with a probe including pH measurement, temperature and automatically compensated for temperature. Freshly cut muscles, about 2 cm thick, were individually placed on white polystyrene foam trays and overwrapped with a transparent polyvinyl chloride (PVC) film permeable to oxygen (12,500 cc/m^2/24h/bar). Afterwards, they were placed in a 4°C refrigerator for 45 min to accommodate for "color blooming" (myoglobin oxygenation) of the samples. The PVC film was removed before color measurement. The tristimulus color values were obtained using the self-constructed CVS as described in Tomasevic et al. (2019) with 10 technical replicates per sample.

2.2. Design of experiment (DoE)

Using the L* a* b* values for pork loin obtained as described above, we choose a central reference value of (57,25,6) in CIELAB space (standard illuminant D65), i.e., L* = 57, a* = 25, b* = 6. Starting there, we moved away in one of 26 defined directions, taking 1,2, …,9 steps of length of ΔE = 1/3. The 26 potential directions were defined by: the L*, a*, and b* axes (plus/minus, which gives 6 directions), all the bisecting lines in L* x a*, L* x b*, and a* x b* subspace (12 directions), and by changing L*, a*, and b* simultaneously and to the same extent (8 directions). In other words, the directions moved away from the reference value are defined by the vectors (L*, a*, b*), L*, a*, b* ∈{−1,0,1}, but excluding (0,0,0); i.e., there are 3^3–1 = 26 possibilities. For illustration, Fig. 1 depicts the levels (circles) and directions (lines) we consider in a* x b* subspace. The resulting a* and b* values are seen at the intersections of circles and lines; here L* remains at the reference value of 57. The circles (levels) give the distance to the reference in L*a*b* space, which means that level 1 (inner circle) corresponds to a ΔE distance of 1/3, level 2 to 2/3, and so on.

The formula used to calculate ΔE between two colors was as follows:

$$\Delta E = \sqrt{\left(L_1^* - L_2^*\right)^2 + \left(a_1^* - a_2^*\right)^2 + \left(b_1^* - b_2^*\right)^2}$$

The resulting design of the experiment (DoE) was balanced with respect to the levels, the directions, and the combination of both, corresponding to 234 colors plus the reference color. Overall, each set (respondent) contained 18 triangle tests (18 different directions, all 9 levels, 2 directions per level), with the concrete level-direction combinations and order of the tests chosen at random. The DoE was calculated across 260 sets (n = 260 respondents).

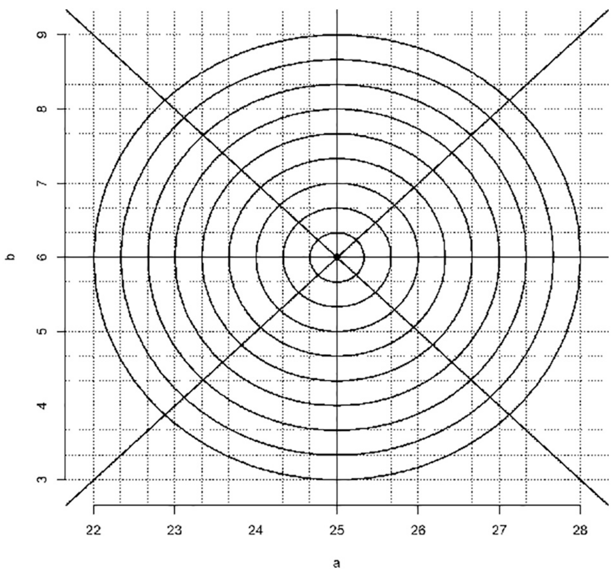

Fig. 1. Depiction of a* x b* subspace used in calculating the 26 defined directions in CIELAB color space.

2.3. Online survey and triangle testing procedure

The online survey was programmed using EyeQuestion sensory research survey software (Logic8 BV, Elst, The Netherlands), with the following components: informed consent, testing for color blindness (deficiency in color perception), triangle tests, sociodemographic questions. A participant had to actively click on a box stating their agreement to participate; if the participant did not agree (clicked on "do not agree"), then the participant was sent to the end screen.

Consenting participants had to first successfully complete a short color deficiency test in order to proceed. Color deficiencies between either turquoise vs. orange, red-brown vs. green, grey vs. red, green vs. ocher, or a deficiency in multiple colors were tested using 5 separate color cards, where a number was visible, if no deficiency was apparent. Respondents were asked to correctly identify the number (multiple-choice question; 4 options) and were given a second chance in the case that an answer was incorrect; their previous answer was not indicated during the re-try.

Next, those respondents who passed the color deficiency test were given information about the triangle test procedure, i.e., to identify the circle that is of a different color from the other two. In order to verify that respondents understood the task, respondents were asked to complete two triangle tests comparing very distinct green and blue circles. Afterwards, the respondents were automatically assigned to the next open set (1 of possible 260), where a respondent proceeded with the DoE predetermined 18 triangle tests corresponding to said set. Each triangle test compared two of 235 possible colors (Fig. 2) and the display order was systematically varied (i.e. AAB, ABA BBA, etc.). The colors were digitalized by converting the L*a*b* values into sRGB color space using convertColor() from package grDevices (version 4.0.2) in R (R Core Team, 2020); see the manual for details. Then the RGB values were input to produce images ("color chips") of the desired colors and exported from R.

After the 18 triangle tests, respondents were asked to overall evaluate the difficulty of the tests using a 7-point hedonic scale ("very easy" to "very difficult"). Finally, demographic questions captured information on age, gender, and profession. The survey was conducted in German.

A question inquiring into the device used for the survey was also included during the sociodemographic section, as the medium of presentation, i.e., screen technology and device, could influence the ability to perceive color differences. In addition, although EyeQuestion is compatible with handheld devices, the program could not resize color chips, so that in case of a very small screen (portrait layout on a smartphone), circles may have overlapped slightly; therefore, making the test slightly easier. For this reason, we asked respondents to identify the description of a device that most closely resembles the device with which they completed the survey. Respondents could select from: iPhone, iPad, Macbook, iMac, (other) smartphone, (other) laptop, (other) tablet, desktop computer with monitor, other. Respondents using smartphones were reminded to hold the screen horizontally throughout the survey to avoid overlapping.

2.4. Respondents and sampling

Using convenience sampling (e.g., university and extracurricular emailing lists, and social networks) a total of 473 individuals accessed the online survey. In the end, 282 complete datasets were collected;

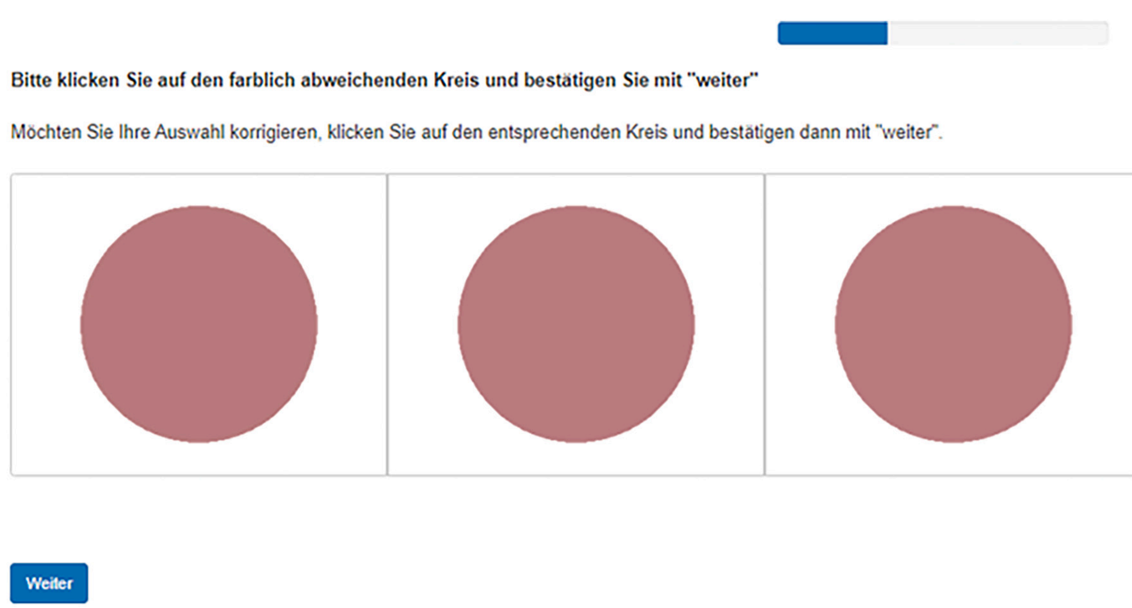

Fig. 2. Example of triangle test as shown in online survey to respondents (survey was conducted in German).

however, the completed number of unique sets was only 254. Therefore, in order to approximately balance the data with respect to the directions (with 174–178 observations for each direction) and level-direction combinations (18–20 observations each) the first completed set was used for analysis, when a set was completed twice. This imbalance was due to software programming constraints: EyeQuestion reassigns previously used sets once all sets are accessed; in combination, completeness of sets needed to be monitored manually. With these constraints, some sets were completed twice and others (6 in total) were excluded due to incompleteness. The overall sample size (number of tests) is 4572 (254 multiplied by 18). Average age of respondents was 29.4 (SD = 11.1) years of age. Respondents identified as female (63.4%), male (35.4%) and diverse (Germany's legal third gender category) (1.2%). Nearly sixty-percent (59.5%) of respondents were students, either attending post- (58.3%) or secondary (high school) education (1.2%) at the time of the survey. The remainder were in the workforce or retired from the workforce. Overall, respondents evaluated the triangle test task as difficult; the average perceived difficulty score was 5.8 (SD = 1.18) on the 7-pt hedonic scale. Furthermore, most respondents used a smartphone (46.9%) to complete the survey; approx. 33% of respondents used an Apple device (e.g., iPhone, iPad, Macbook, or iMac).

2.5. Data analysis

Analyses were estimated using R statistical software (Version 4.0.2) (R Core Team, 2020), while applying add-on packages lme4 (Bates, Maechler, Bolker, & Walker, 2015) and plot3D (Soetaert, 2019). The package sensR (Christensen & Brockhoff, 2020) was used to calculate *d'* values. Data, R code, and supporting information to reproduce our analysis are available at Zenodo.org (Gertheiss et al., 2021).

When analyzing the data obtained, we started by modeling the binary response of correct identification (yes/no) using a generalized linear logistic mixed model with predictor ΔE (fixed effect) and subject-specific random slope. Instead of an intercept that is estimated from the data, we include an offset (of $log(p_0/(1-p_0))$) yielding the probability of success $p_0=1/3$ for pure guessing if $\Delta E \rightarrow 0$. Results indicated a positive and highly significant effect of ΔE ($P < 0.001$).

Next, we relaxed the assumption of linearity (in ΔE) in order to identify the first level (according to ΔE values considered) with a success rate that is significantly higher than guessing; we replaced the linear effect of ΔE by a factor giving the 9 considered levels (i.e., steps in ΔE by 1/3).

In order to investigate the effect sizes of different ΔE values and to compare the linear and the factor models, we plot the effects as a function of ΔE (Fig. 3). Based on Fig. 3, it appears that the effect of ΔE is non-linear. This is also confirmed by a likelihood ratio test comparing the two models ($P = 0.037$).

Finally, as another alternative to the linear model, we also considered a purely quadratic model. The reason to choose a purely quadratic modeling is that a zero effect, with a vanishing gradient, can be obtained for $\Delta E \rightarrow 0$. The quadratic model confirms the significant effect of ΔE ($P < 0.001$), as with the linear model. Furthermore, the quadratic model estimates the effect of ΔE similar to the factor model; no significant improvement is found when comparing the factor model to the quadratic model (Fig. 4). As a consequence, we chose the quadratic model as a good compromise for sparsity and interpretability/plausibility, as well as fit to the data.

Applying the quadratic model for further analysis, we estimated the probability of correctly answering a triangle test based on the ΔE level. We also checked for potential effects of covariates (e.g., age, gender, device, occupation, and perceived difficulty); no significant covariates were identified (which in case of age might be explained by the fact that the sample was not diverse enough, compare *Respondents and sampling* above). In order to estimate the effect of direction, and test our second hypothesis, we (i) let the effect of $(\Delta E)^2$ vary with each direction considered (i.e., according to the 26 DoE-pre-determined directions) in a quadratic model. To distinguish the effects of L*, a*, and b*, we (ii) allowed for different effects of $(\Delta L^*)^2$, $(\Delta a^*)^2$, $(\Delta b^*)^2$ in order to best ascertain effect sizes based on the three dimensions. In the latter case (ii), the initial model with quadratic effect of ΔE can be obtained by setting two linear restrictions, i.e., the effects of the dimensions are all considered equal. Both the model (i) with 26 different effects of $(\Delta E)^2$ and the model (ii) with different effects of $(\Delta L^*)^2$, $(\Delta a^*)^2$, $(\Delta b^*)^2$ led to an improvement over the simple model with $(\Delta E)^2$ only, as shown by likelihood ratio tests ((i) $P \approx 0.001$ and (ii) $P \approx 6 \times 10–5$, respectively).

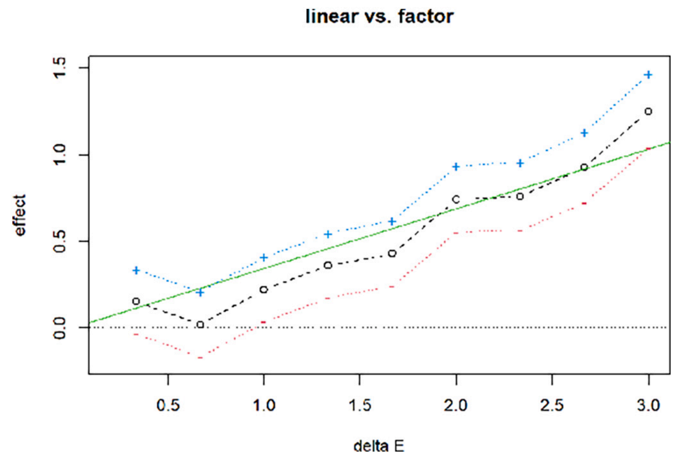

Fig. 3. ΔE effect size as estimated by a linear or factor model. Black circles indicate fitted effects of the factor model ±2 std. errors (blue vs. red), i.e., approximate pointwise 95% confidence intervals. The green line signifies the linear model estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

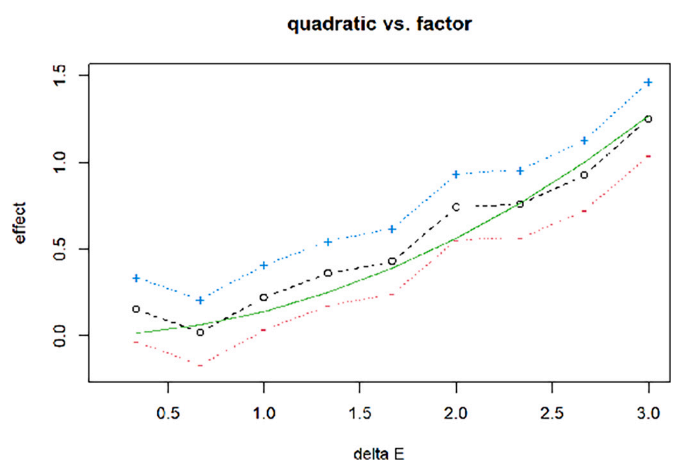

Fig. 4. ΔE effect size as estimated by a factor or quadratic model. Black circles indicate fitted effects of the factor model ±2 std. errors (blue vs. red), i.e., approximate pointwise 95% confidence intervals. The green line signifies the quadratic model estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3. Results & discussion

3.1. Effect of ΔE on color discrimination

Initially, the triangle tests were evaluated according to the proportion of correct answers (Table 1). Already on the basis of these descriptive results it becomes clear that color discrimination becomes easier with an increasing ΔE value. As 33% of correct answers would be the proportion left to chance, i.e., in a triangle test a respondent has a 1:2

Table 1
Frequency count of correctly answered triangle tests and *d'*-value corresponding to specific ΔE levels applied in DoE.

Level	Associated ΔE (*n* = 508)	Frequency of correct answers	Percent correct	*d'* [95% CI]
1	1/3	186	36.61%	0.61 [0.00, 0.95]
2	2/3	172	33.85%	0.24 [0.00, 0.74]
3	1	196	38.58%	0.77 [0.33, 1.07]
4	1 1/3	212	41.73%	1.00 [0.68, 1.26]
5	1 2/3	221	43.50%	1.11 [0.82, 1.36]
6	2	259	50.98%	1.52 [1.28, 1,74]
7	2 1/3	261	51.38%	1.54 [1.30, 1.76]
8	2 2/3	280	55.12%	1.73 [1.50, 1.95]
9	3	315	62.01%	2.08 [1.86, 2.30]
	Total*	2102	45.98%	

CI corresponds to 95% confidence interval
* $n = 4572$, corresponding to 508 observations for each of the 9 levels.

chance of guessing correctly, we see that ΔE = 1 (around 40% correct answers) could be a threshold at which color differences become clearly discernable to consumers. This is supported by the found *d'* value (1.00) at ΔE = 1 which is considered a threshold for discrimination (Ishii, Kawaguchi, O'Mahony, & Rousseau, 2007). At a ΔE = 3, more than 60% of respondents were able to correctly discriminate between the presented colors within the triangle tests. Although this points to color discrimination, generally, it depicts that even a ΔE = 3 may not always be relevant (or discernable) for all consumers in the realm of perceiving differences in meat color.

However, descriptive frequencies are not enough to test the two stated hypotheses, nor to test the effect size and statistical significance of our results. The frequencies stated in Table 1 also do not take differing respondent abilities into account. Therefore, the triangle tests were evaluated using a series of generalized logistic mixed models, as described in section *Data Analysis* above.

All 3 generalized logistic mixed models verify an effect of ΔE on a person's ability to discern between two colors. Specifically, the factor model points to changes in colors being first detectable at ΔE =1 (Fig. 4), confirming the supposition made based on Table 1. Significance as well as coefficient estimates increase steadily from that point onwards (Table 2). The high significance of coefficients in the factor model clearly depicts that a ΔE > 1 results in respondents, on average, being able to discern differences in colors. Nonetheless, it still remains unclear to what extent this value is relevant across a large population, as significance does not automatically equate to relevance.

As expressed in Section 2.5, we considered a quadratic model as best able to model the effect of ΔE on color discrimination due to the smoothing of estimates and therefore ease of interpretability. The random (subject-specific) effect's standard deviation is large (SD = 0.2416), meaning that there is substantial variation in respondents' abilities to differentiate colors. This explains why even at larger ΔE values (> 1), the frequency of correct answers still does not approach 100%.

In order to explore respondents color differentiation abilities, we estimated probabilities of correctly answering a triangle test based on the quadratic model (Fig. 5). The quadratic model estimated that the mid segment of 68% respondents can correctly differentiate colors with a ΔE = 3 between 47% and 78% of the time. The most adept (top 16%) are at least 78% likely to correctly differentiate the colors at ΔE = 3; the bottom 16% are only likely to identify color differences less than 47% of the time. In Fig. 5, we still see that ΔE = 1 is a critical value where the probability of correctly answering is distinctly above the probability of guessing for about 50% of the respondents. The bottom 16% "poor performers" eclipse the 33% probability of guessing correctly at approx. ΔE = 1.8.

In other words, the blue plus curve (and above) indicates the top 16% of respondents. The black curve represents the median of respondents. The red curve with minus symbols (and below) indicates the bottom 16% of respondents. Further note, the black circle is not the middle between + and − due to the logit link.

Overall, based on our results we must reject our first hypothesis. Our results indicate that a ΔE < 2, primarily a ΔE = 1, may indeed be a critical value when assessing the general public's ability to ascertain color differences in monochrome products. This reports a much lower level than the range which has been previously assumed in the meat science literature (Tomasevic et al., 2019) and should be taken into consideration while planning and evaluating color research in meat science. In addition, we do not see a significant effect of age on color differentiation, there is a trend (i.e., probability of a correct answer decreases) according to age. The covariate likely remains insignificant, as the proportion of older respondents is relatively small in our sample. Furthermore, device used to complete the survey did not significantly affect the results, meaning that we can assume most respondents completed the exercise without overlapping colors, i.e. smartphones held horizontally.

Table 2
Coefficients for differing ΔE levels as estimated by the factor model (based on generalized linear logistic mixed model).

Level	Associated ΔE	Coefficient	Std. Error	*P*-value	Significance
1	1/3	0.1520	0.0922	0.099	
2	2/3	0.0195	0.0946	0.837	
3	1	0.2218	0.0930	0.017	*
4	1 1/3	0.3594	0.0932	0.0001	***
5	1 2/3	0.4300	0.0945	5.35 × 10–06	***
6	2	0.7429	0.0959	9.56 × 10–15	***
7	2 1/3	0.7607	0.0984	1.06 × 10–14	***
8	2 2/3	0.9270	0.1016	< 2 × 10–16	***
9	3	1.2507	0.1069	< 2 × 10–16	***

Significance codes: '***' $P < 0.001$. '**' $P < 0.01$. '*' $P < 0.05$

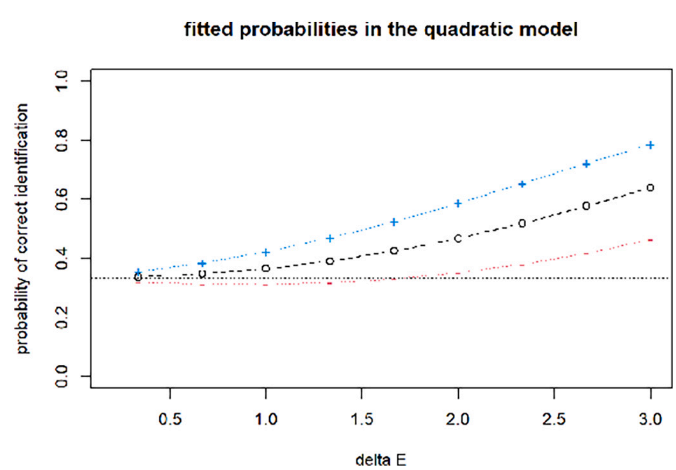

Fig. 5. Estimated/fitted probability of differentiating colors correctly based on ΔE. The black curve indicates the probability for a zero random effect, the blue/red curve indicates +/− one standard deviation of the random/subject-specific effect. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Effect of dimension and direction on color discrimination

When using the relaxed quadratic model (ii) as explained in *Data Analysis*, we see that the coefficient (on the logit scale) of $(\Delta b^*)^2$ is about twice the intensity of the other 2 dimensions; although all 3 dimensions are found to have a significant effect on color discrimination (Table 3).

Although we checked and there was no indication that the effect of L*, a*, or b* depended on the value relative to the reference midpoint color; we do see that the overall effect of ΔE may indeed change depending on the direction (26 axes) we choose to move away from the reference in L*a*b* space. In order to visualize our results in the most flexible model (i) with different effects for each of the 26 directions considered, a 3D model was created based on the 26 axes from the DoE. Fig. 6 shows the effect of ΔE for each direction separately, as modelled with a quadratic transformation of data. Here, it is apparent that the axes in the b* dimension influence discrimination the most, particularly along with the a* x b* axes. L* appears to have the least impact on color discrimination.

Our findings mirror the observations of Zhu and Brewer (1999) (Zhu & Brewer, 1999) in that both a* and b* dimensions appear play a role in the perception of meat redness. This could likely be due to these dimensions having a specifically intense influence on the ability of a person to discern different color hues in the spectrum of meat color, as proven through this study. Particularly, our results point to the b* dimension to having a larger influence in perceiving meat color than originally thought.

4. Implications & limitations

The low ΔE values (a ΔE between 1 and 2) necessary to discern colors supports the need for product sorting and color monitoring as a necessary meat product quality criterion, not only from a biophysical standpoint (Joo, Kauffman, Kim, & Park, 1999; Kim et al., 2010; Mancini & Hunt, 2005), but also from an aesthetic standpoint of how consumers assess meat quality (consumer perception) (Bello Acebrón & Dopico, 2000; Kennedy et al., 2004; Kennedy, Stewart-Knox, Mitchell, & Thurnham, 2005; Ngapo et al., 2018). The findings of this study are fundamental in moving consumer preference and acceptance research forward regarding product color.

Product color is frequently listed as one of multiple (Bello Acebrón & Dopico, 2000; Droval et al., 2012; Kennedy et al., 2004; Ngapo et al., 2018) search criteria for raw meat. And currently, although there are technologies to capture reliable color measurements in meat (Mancini & Hunt, 2005; Tomasevic et al., 2019), consumer preference research has focused on comparisons of subjectively described (e.g., Bello Acebrón & Dopico, 2000; Mahbubi, Uchiyama, & Hatanaka, 2019) or non-systematically chosen “red” vs. “pink” vs. “white” hues (e.g., Ngapo et al., 2007). Our clear indication of a ΔE at which colors are discernable provides the required information to carry-out more systematic investigations of consumer preferences for meat color. Furthermore, the findings of future studies can help industry to supply products to the associated appropriate markets and consumers with a specific preference.

Our results also add to the discussion on the proper techniques to measure the color of meat and how to determine relevant differences in meat color. Tomasevic et al. (2019) have already shown that traditional techniques, such as a spectrophotometer, do not provide reliable color measurements for meat. This has two implications in association with this study. First, variation in color measurements using a spectrophotometer could confound color differences perceptible to the human eye. Secondly, spectrophotometers inaccurately represent visually perceived colors when reproducing images based on spectrophotometer CIE L*a*b* values (Tomasevic et al., 2019). Panelists found CVS reproduced colors to be more similar to the actual product than colors reproduced using spectrophotometer measurements (Tomasevic et al., 2019. This further supports the need for CVS data for planning studies to determine consumer preferences of meat color. Furthermore, our research demonstrates that quantitatively small differences in color are perceivable to the human eye and could be of relevance when sorting meat products. Future research has yet to ascertain the ΔE threshold of acceptance for visually dissimilar meat products; this is also critical to answer when it comes to efficiently packaging meat for the end consumer.

Despite the central findings, which contribute to the further investigation of color perception and consumer perception of meat quality, we must admit to some limitations. Firstly, our DoE is founded on the CVS color measurements of only 15 pork loin (*Longissimus thoracis et lumborum*) samples. This certainly does not represent the true variation in pork meat color available on the global market and influences the centre of the DoE. Future research should investigate the repeatability of our experiment using different meat color ranges; i.e. the experiment should be repeated with a shifted center of the DoE.

Secondly, the high variability of color discrimination encountered in our experiment is likely to have been influenced by the medium on which respondents completed the survey. The online survey format left numerous possibilities for respondents to encounter different testing situations. For example, the largest group of respondents completed the survey on a smartphone. Although EyeQuestion is equipped to modify the survey format to accommodate smartphone software, the software was not able to resize the color chip images, meaning that if a respondent chose to complete the survey on a vertical smartphone screen the colors in the triangle tests overlapped slightly. To allay this effect, respondents were asked and reminded to hold the screen horizontally for the entirety of the survey. Differing screen brightness and display technology of respondents’ screen may have also influenced color discrimination. The experiment was initially planned at a central location with color-calibrated display screens for April 2020. Due to the global pandemic and recurring national restrictions the study was transferred to a de-centralized online format. Therefore, this experiment should be replicated in a central location with uniform and calibrated equipment in order to verify the results reported here.

Despite the uncertainties encountered using an online format, we are still confident in our results as a starting point for future systematic research into consumer perception and preferences for meat color, and human color perception in general. We did not find the device to be a

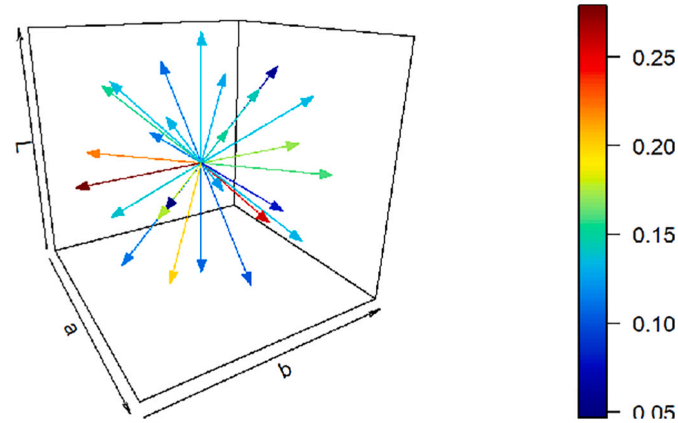

Fig. 6. The effect of $(\Delta E)^2$ corresponding to direction within L*a*b* space. Arrow color denotes effect size.

Table 3
Effect of L*a*b* color space dimensions on color discrimination.

Associated dimension	Coefficient	Std. Error	*P*-value	Significance
L*	0.0994	0.0190	1.61 × 10–7	***
a*	0.1087	0.0191	1.29 × 10–8	***
b*	0.2162	0.0198	< 2 × 10–16	***

Significance codes: ‘***’ $P < 0.001$. ‘**’ $P < 0.01$. ‘*’ $P < 0.05$

statistically significant covariate in our models. More importantly, our results lie within a ΔE range found to be relevant for white hues in dentistry (ΔE ≈ 1.8) (Thoma et al., 2016) as well as imaging and color sciences (ΔE ≈ 2) (Pointer et al., 2002a, 2002b; Stokes et al., 1992), illustrating that differences in display specifications may not be of great importance when it comes to testing color discrimination. However, these results should be repeated with a central location test to validate online testing with a balanced design of experiment (DoE) and a large enough respondent sample size.

A third limitation of our study is the rather young respondent age. Medical research points to decreasing color perception with increasing age (Smith & Pokorny, 2003). Unfortunately, due to our sampling techniques and resulting sample composition we did not observe a significant effect of age on color discrimination. Even so, the color deficiency testing employed as a screening procedure may have excluded individuals who would have increased the ΔE value found to be critical for color differentiation. This begs the question whether such a screening criterion should be employed when trying to identify a threshold or relevant value for the population at large.

Conclusively, we can say that our research advances the fundamental understanding of consumers' color perception, particularly in the realm of meat science, and that these findings are foundational to improve further research in consumer preferences, not only for meat color. Furthermore, this study showcases an efficient and innovative systematic method to test consumers' ability to differentiate between colors, which could be expanded to other research areas on product appearance and color. Using this computer-based discrimination testing approach, we show that large scale evaluation of human color perception is possible.

Funding sources

This work was supported by the German Federal Ministry of Education and Research (Nr. 01PL16061).

Author statement

Brianne A. Altmann[a]*: study conceptualization, data collection and analysis, manuscript writing & editing, discussion of results and implications.

Jan Gertheiss[b]: study design and statistics, manuscript writing (Material and methods) and editing.

Igor Tomasevic[c]: study conceptualization, meat color data collection (CVS measurements), manuscript writing (material and methods), discussion of results and implications, manuscript editing.

Christina Engelkes[a], Thibaud Glaesener[a], Jule Meyer[a], Alina Schäfer[a], Richard Wiesen[a]: study conceptualization, data collection and analysis, manuscript writing, discussion of results and implications.

Daniel Mörlein[a]: study conceptualization, discussion of results and implications, manuscript editing.

Declaration of Competing Interest

None.

Acknowledgements

This research was carried out as part of the Research-Oriented Teaching and Learning (*Forschungsorientiertes Lehren und Lernen*) program funded by the German Federal Ministry of Education and Research (Funding Nr.: 01PL16061). Program and teaching organizational support at the University of Goettingen was provided by Susanne Wimmelmann. Elisa Oertel assisted in data visualization.

References

Adzitey, F., & Nurul, H. (2011). Pale soft exudative (PSE) and dark firm dry (DFD) meats: Causes and measures to reduce these incidences - A mini review. *International Food Research Journal, 18*(1), 11–20.

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

Bello Acebrón, L. B., & Dopico, D. C. (2000). The importance of intrinsic and extrinsic cues to expected and experienced quality: an empirical application for beef. *Food Quality and Preference, 11*(3), 229–238. https://doi.org/10.1016/S0950-3293(99)00059-2

Chen, C. C., Wardman, R. H., & Smith, K. J. (2002). The mapping of a surface of constant visual depth in CIELAB colour space. *Coloration Technology, 118*(6), 281–294. https://doi.org/10.1111/j.1478-4408.2002.tb00112.x

Christensen, R., & Brockhoff, P. (2020). sensR—An R-package for sensory discrimination. Retrieved from http://www.cran.r-project.org/package=sensR/.

Droval, A. A., Benassi, V. T. T., Rossa, A., Prudencio, S. H. H., Paiao, F. G., Shimokomaki, M., & Shimokomaki, M. (2012). Consumer attitudes and preferences regarding pale, soft, and exudative broiler breast meat. *Journal of Applied Poultry Research, 21*(3), 502–507. https://doi.org/10.3382/japr.2011-00392

Gertheiss, J., Engelkes, C., Glaesener, T., Meyer, J., Schäfer, A., Wiesen, R., … Mörlein, D. (2021). *235 shades of pink*. https://doi.org/10.5281/zenodo.4683190

Ishii, R., Kawaguchi, H., O'Mahony, M., & Rousseau, B. (2007). Relating consumer and trained panels' discriminative sensitivities using vanilla flavored ice cream as a medium. *Food Quality and Preference, 18*(1), 89–96. https://doi.org/10.1016/j.foodqual.2005.08.004

Joo, S. T., Kauffman, R. G., Kim, B. C., & Park, G. B. (1999). The relationship of sarcoplasmic and myofibrillar protein solubility to colour and water-holding capacity in porcine longissimus muscle. *Meat Science, 52*(3), 291–297. https://doi.org/10.1016/S0309-1740(99)00005-4

Kennedy, O. B., Stewart-Knox, B. J., Mitchell, P. C., & Thurnham, D. I. (2004). Consumer perceptions of poultry meat: A qualitative analysis. *Nutrition & Food Science, 34*(3), 122–129. https://doi.org/10.1108/00346650410536746

Kennedy, O. B., Stewart-Knox, B. J., Mitchell, P. C., & Thurnham, D. I. (2005). Flesh colour dominates consumer preference for chicken. *Appetite, 44*(2), 181–186. https://doi.org/10.1016/j.appet.2004.11.002

Kim, G., Jeong, J., Hur, S., Yang, H., Jeon, J., & Joo, S. (2010). *The relationship between meat color (CIE L * and a *), myoglobin content, and their influence on muscle fiber characteristics and pork* quality. *30*(4), 626–633.

La Rosa, G. R. M., Pasquale, S., Pedullà, E., Palermo, F., Rapisarda, E., & Gueli, A. M. (2019). Colorimetric study about the stratification's effect on colour perception of resin composites. *Odontology*. https://doi.org/10.1007/s10266-019-00469-9, 0123456789.

Liu, Q., Huang, Z., Pointer, M. R., Luo, M. R., Xiao, K., & Westland, S. (2018). Evaluating colour preference of lighting with an empty light booth. *Lighting Research and Technology, 50*(8), 1249–1256. https://doi.org/10.1177/1477153517727330

Lusk, J. L., Tonsor, G. T., Schroeder, T. C., & Hayes, D. J. (2018). Effect of government quality grade labels on consumer demand for pork chops in the short and long run. *Food Policy, 77*, 91–102. https://doi.org/10.1016/j.foodpol.2018.04.011

Mahbubi, A., Uchiyama, T., & Hatanaka, K. (2019). Capturing consumer value and clustering customer preferences in the Indonesian halal beef market. *Meat Science, 156*(April), 23–32. https://doi.org/10.1016/j.meatsci.2019.05.012

Mancini, R. A., & Hunt, M. C. (2005). Current research in meat color. *Meat Science, 71*(1), 100–121. https://doi.org/10.1016/j.meatsci.2005.03.003

McMillin, K. W. (2008). Where is MAP going? A review and future potential of modified atmosphere packaging for meat. *Meat Science, 80*(1), 43–65. https://doi.org/10.1016/j.meatsci.2008.05.028

Mörlein, D., Link, G., Werner, C., & Wicke, M. (2007). Suitability of three commercially produced pig breeds in Germany for a meat quality program with emphasis on drip loss and eating quality. *Meat Science, 77*(4), 504–511. https://doi.org/10.1016/j.meatsci.2007.04.030

Ngapo, T. M., Martin, J. F., & Dransfield, E. (2007). International preferences for pork appearance: II. Factors influencing consumer choice. *Food Quality and Preference, 18*(1), 139–151. https://doi.org/10.1016/j.foodqual.2005.09.007

Ngapo, T. M., Riendeau, L., Laberge, C., & Fortin, J. (2012). "Chilled" pork - Part II. Consumer perception of sensory quality. *Meat Science, 92*(4), 338–345. https://doi.org/10.1016/j.meatsci.2012.04.031

Ngapo, T. M., Rubio Lozano, M. S., & Braña Varela, D. (2018). Mexican consumers at the point of meat purchase. Pork choice. *Meat Science, 135*(August 2017), 27–35. https://doi.org/10.1016/j.meatsci.2017.08.005

Pointer, M. R., Attridge, G. G., & Jacobson, R. E. (2002a). Food colour appearance judged using images on a computer display. *Imaging Science Journal, 50*(1), 23–36. https://doi.org/10.1080/13682199.2002.11784389

Pointer, M. R., Attridge, G. G., & Jacobson, R. E. (2002b). Perceived colour differences in displayed colours part 1: Hard copy to soft copy matching. *Imaging Science Journal, 50*(1), 1–9. https://doi.org/10.1080/13682199.2002.11784387

R Core Team. (2020). R: A language and environment for statistical computing. Retrieved from https://www.r-project.org/.

Smith, V. C., & Pokorny, J. (2003). Color matching and color discrimination. In S. K. Shevell (Ed.), *The science of color* (2nd ed., pp. 103–148). Oxford, UK: Optical Society of America & Elsevier.

Soetaert, K. (2019). plot3D: Plotting multi-dimensional data. Retrieved from https://cran.r-project.org/package=plot3D.

Stokes, M., Fairchild, M. D., & Berns, R. S. (1992). Precision requirements for digital color reproduction. *ACM Transactions on Graphics, 11*(4), 406–422. https://doi.org/10.1145/146443.146482

Tapp, W. N., Yancey, J. W. S., & Apple, J. K. (2011). How is the instrumental color of meat measured? *Meat Science, 89*(1), 1–5. https://doi.org/10.1016/j.meatsci.2010.11.021

Thoma, D., Ioannidis, A., Fehmer, V., Michelotti, G., Jung, R., & Sailer, I. (2016). Threshold values for the perception of color changes in human teeth. *The International Journal of Periodontics & Restorative Dentistry, 36*(6), 777–783. https://doi.org/10.11607/prd.2937

Tomasevic, I., Djekic, I., Font-i-Furnols, M., Terjung, N., & Lorenzo, J. M. (2021). Recent advances in meat color research. *Current Opinion in Food Science, 41*, 81–87. https://doi.org/10.1016/j.cofs.2021.02.012

Tomasevic, I., Tomovic, V., Milovanovic, B., Lorenzo, J., Đorđević, V., Karabasil, N., & Djekic, I. (2019). Comparison of a computer vision system vs. traditional colorimeter for color evaluation of meat products with various physical properties. *Meat Science, 148*, 5–12. https://doi.org/10.1016/j.meatsci.2018.09.015

Zhu, L. G., & Brewer, M. S. (1999). Relationship between instrumental and visual color in a raw, fresh beef and chicken model system. *Journal of Muscle Foods, 10*(2), 131–146. https://doi.org/10.1111/j.1745-4573.1999.tb00391.x