Identifying drivers of liking and characterizing the ideal product thanks to Free-Comment

Benjamin Mahieu, Michel Visalli, Pascal Schlich

Version of Record: https://www.sciencedirect.com/science/article/pii/S0950329321002718
Manuscript_aa13621f0c7dcf98f610584e86e4ca41

Title

Identifying drivers of liking and characterizing the ideal product thanks to Free-Comment

Authors

Benjamin Mahieu[1,2], Michel Visalli[1,2], Pascal Schlich[1,2]

[1]Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE, Université Bourgogne Franche-Comté, F-21000 Dijon, France.

[2]INRAE, PROBE research infrastructure, ChemoSens facility, F-21000 Dijon, France

Corresponding author:

Benjamin Mahieu; Pascal Schlich

benjamin.mahieu@inrae.fr; pascal.schlich@inrae.fr

Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE, Université Bourgogne Franche-Comté, F-21000 Dijon, France.

INRAE, PROBE research infrastructure, ChemoSens facility, F-21000 Dijon, France

Highlights

- Drivers of liking can be identified from Free-Comment (FC) data
- Ideal-Free-Comment (IFC) enables to characterize the ideal product with FC
- IFC enables to locate the ideal product on the FC product map
- Drivers of liking based on FC data and IFC are complementary tools

Abstract

Consumers' hedonic appreciation is important for the commercial success of a product. To formulate appreciated products, sensory and hedonic data of some existing products are often linked to each other. Because existing products represent only a limited sensory space of investigation, asking consumers to characterize their ideal product can provide relevant additional information to understand their preferences. First, the paper investigates whether sensory drivers of liking can be derived from linking Free-Comment (FC) and hedonic data. Second, Ideal-Free-Comment (IFC) is introduced. IFC instructs consumers to describe actual products and then their ideal product thanks to FC. IFC paired with liking scoring was used in a home-use test with 483 consumers each evaluating from 1 to 14 (5.71 on average) cooked hams from a list of 30 hams representative of the French market. Based on a mixed linear model, relevant drivers of liking were identified from FC data. The panel's average ideal product was consistent with the drivers of liking. Since descriptors with opposite meanings characterized individual ideal products, a consumer segmentation based on their ideal product was performed and resulted in two segments. The two segments' ideal products mainly differed regarding their flavor. Drivers of liking and the ideal product of the smaller segment (≈ 15% of the consumers) were not well consistent suggesting this was a noise segment. Drivers of liking based on FC data and IFC are complementary tools to understand consumers' hedonic appreciation without the use of a pre-established list of descriptors.

Keywords

- Open-ended questions
- Drivers of liking
- Ideal-Free-Comment (IFC)
- Consumer segmentation
- Cooked ham
- Home-Use Test (HUT)

1. Introduction

Consumers' hedonic appreciation is one of the most important drivers of the commercial success of a product. It is most often investigated using hedonic tests in which a panel of consumers is instructed to score their overall liking of products. Since liking is a function of the products' sensory characteristics (Lagrange & Norback, 1987), investigating these characteristics is necessary to understand liking and formulate appreciated products. For this reason, hedonic tests are often performed conjointly to the sensory characterization of the products. Because consumers were claimed not to be able to provide valid nor reliable sensory characterization (Ares & Varela, 2017; Lawless & Heymann, 1999; Meilgaard, Civille, & Carr, 1991; Stone & Sidel, 1993), this sensory characterization used to be performed by sensory profiling using a trained panel.

Several methodologies have been developed to link sensory and hedonic data among which preference mapping techniques (Carroll, 1972; Danzart, 2009; Greenhoff & MacFie, 1994; McEwan, 1996; Schlich & McEwan, 1992) are likely the most popular. Two major approaches can be distinguished among preference mapping techniques: internal preference mapping and external preference mapping. They mainly differ in the point of view they adopt (van Kleef, van Trijp, & Luning, 2006). Internal preference mapping puts the focus on the hedonic data: the product space is obtained from liking scores and the sensory descriptor scores are regressed into this space. On the contrary, external preference mapping puts the focus on the sensory data: the product space is obtained from sensory descriptor scores and the individual liking scores are regressed into this space. Worch (2013) proposed the so-called prefMFA method that uses Multiple Factor Analysis (Escofier & Pagès, 1994) to determine the shared dimensions between sensory and hedonic data.

During the last recent years, the affirmation upon which consumers are unable to provide valid or reliable sensory characterization has been reconsidered. One of the main reasons is that trained panels might consider descriptors and variations that are irrelevant to consumers (Ares & Varela, 2017; ten Kleij & Musters, 2003). In addition, several consumer methods were claimed to obtain more or less similar information as the one provided by sensory profiling in practical applications (Ares & Varela, 2017; Valentin, Chollet, Lelièvre, & Abdi, 2012; Varela & Ares, 2012). Among these consumer methods, some were specifically designed to understand preferences and to link hedonic data with consumer sensory data. Notably, Just-About-Right (JAR) scales (see for example Popper (2014)) and Check-All-That-Apply (CATA) (Adams, Williams, Lancaster, & Foley, 2007) paired with hedonic data collection and penalty-lift analysis (Meyners, Castura, & Carr, 2013) belong to these methods. Other methods sharing the same objective can be mentioned such as Preferred Attribute Elicitation (Grygorczyk, Lesschaeve, Corredig, & Duizer, 2013), preference mapping based on Sorting (Faye et al., 2006), and preference mapping based on CATA (Dooley, Lee, & Meullenet, 2010).

Previous methodologies intend to understand the sensory characteristics that drive the liking and the disliking of the products through the study of some existing products, which necessarily restricts the sensory space investigated. This limitation can affect the conclusions drawn since the ideal product does not necessarily lie within the product space (van Trijp, Punter, Mickartz, & Kruithof, 2007). Indeed, since only a limited number of products are presented to consumers then only a limited number of combinations of sensory characteristics are represented and evaluated. To circumvent this limitation, the Ideal-Profile-Method (IPM) (Moskowitz, 1972; van Trijp et al., 2007; Worch, Lê, Punter, & Pagès, 2013) was proposed. In IPM, consumers are instructed to rate the products on several descriptors from a pre-established list using intensity scales. Right after the evaluation of every actual product, consumers are instructed to do the same task but considering a virtual ideal product. The idea is that the consumers provide for each descriptor the rating they would have found ideal in the previous actual product. Recently, characterizing the ideal product like the actual products has been successfully extended to other methodologies than intensity scales such as CATA (Ares, Dauber, Fernández, Giménez, & Varela, 2014; Ares et al., 2017; Ares, Varela, Rado, & Giménez, 2011; Bruzzone et al., 2015), Projective Mapping (Ares et al., 2011) and Pairwise Comparison (Brard & Lê, 2016). These studies suggest that characterizing the ideal product is relevant even when it is not performed using intensity measurements of each descriptor and when only a single ideal product is considered for each consumer.

Until now, most of the existing methodologies that aim at investigating drivers of liking and characterizing the ideal product are based on a pre-established list of descriptors, which comes with several limitations. The list is tedious to establish and represents a critical aspect for the relevance of the collected data as it may affect the results of the study (Ares et al., 2013). The list may raise consumers' awareness on descriptors they would not think about otherwise (Coulon-Leroy, Symoneaux, Lawrence, Mehinagic, & Maitre, 2017; Kim, Hopkinson, van Hout, & Lee, 2017; Krosnick, 1999). Since the list contains only a limited number of descriptors, it could result in a loss of information and the collected data can be biased by the dumping effect (Krosnick, 1999; Varela et al., 2018). When used in a CATA task, the list likely leads to an acquiescence bias (Callegaro, Murakami, Tepman, & Henderson, 2015;

Kim et al., 2017; Krosnick, 1999), which encourages consumers to check the proposed descriptors.

Luc, Lê, and Philippe (2020) took a step forward in the characterization of the ideal product without the use of a pre-established list of descriptors by proposing the so-called Free JAR profiling. In Free JAR profiling, consumers are instructed to describe a set of products using free descriptions constrained to a JAR syntax. In Free JAR profiling, the ideal product is not directly characterized since its characteristics are derived from the Free JAR descriptions of the actual products. This can result in some loss and/or some misleading information regarding the ideal product if the actual products are not carefully chosen.

Free-Comment (FC) (ten Kleij & Musters, 2003), where consumers are instructed to describe the products using their own terms into free descriptions without syntax constraint, appears as a natural alternative to identify drivers of liking and to characterize the ideal product avoiding the limitations from the existing methodologies. Accordingly, first, the present paper investigates the relevance of FC sensory data to be linked to hedonic data with the final aim of identifying drivers of liking. Second, the Ideal-Free-Comment (IFC) method is introduced and its ability to provide a relevant characterization of the ideal product is investigated. In IFC, consumers are instructed to describe actual products and then their ideal product thanks to FC. In comparison to Free JAR profiling, IFC renders the characterization of the ideal product as independent as possible from the characterization of the actual products with the same benefit of not restricting the sensory characterizations to a pre-established list of descriptors. The final objective was to investigate whether drivers of liking and the ideal product provide consistent, and eventually complementary, information.

2. Material and methods

2.1. Participants

483 consumers from 7 French cities (Agen, Angers, Bourg en Bresse, Caen, Dijon, La Rochelle, Strasbourg) were recruited by technical centers from the ACTIA network and by the SensoStat Company. Among these consumers, 58% were females, 19% were between 18 and 30 years old, 47% were between 31 and 51 years old and 34%

were more than 51 years old. They were selected as being consumers of cooked ham at least once every two weeks and were informed that they should purchase and evaluate a minimum of 4 different hams among a provided list of 30 hams widely available on the French market. Compensation for their participation was 2.5 € for each different evaluated product and no additional compensation was given to those who evaluated more than 12 different products.

2.2. Products

A list of 30 cooked hams of the French Market was selected to span the variability of fat and salt contents observed in this market. This sample was restricted to hams without rind and excluded smoked, braised, spit-roasted, and flavored hams.

2.3. Data acquisition

2.3.1. General procedure

The consumers purchased the products they evaluated and performed the evaluations at home. Each product they evaluated had to be one of the 30 products belonging to the proposed list. An email was sent to the consumers to invite them to connect to TimeSens© (INRAE, Dijon, France) each time they evaluated a product. At each connection, the consumers had to type the European Article Numbering (*EAN*) of the ham they purchased. The consumers could not start the evaluation of a product they already evaluated as this was verified thanks to the EAN. To ensure they bought the product, they had to take a picture of the package, before and after opening. The study lasted 13 weeks and consumers could purchase hams whenever they decided but they were restricted to a maximum of one evaluation per day. Despite consumers were instructed to evaluate a minimum of 4 different hams and were compensated up to 12 ones, some of them evaluated less than 4 and others more than 12. Consequently, the number of hams evaluated by each consumer actually ranged between 1 and 14 (mean = 5.71, sd = 2.47) resulting in a total of 2758 evaluations. The data from consumers not respecting instructions were kept, as every information is good to take. The number of evaluations by ham ranged between 8 and 263 (mean = 91.93, sd = 63.38).

2.3.2. Sensory and hedonic characterization of the actual products

For each evaluated product, it was recalled to consumers to evaluate and consume the product on its own without extra food. They first performed an FC task by sensory modality in the following order: visual aspect, texture in mouth, and flavor. For each sensory modality, the following instructions were given to the consumers:

- Visual aspect: “Please describe the visual aspect of this ham”
- Texture in mouth: “Please describe the texture in mouth of this ham”
- Flavor: “Please describe the taste of this ham”

Right after the FC task, the consumers rated their liking of the product using a 0-10 VAS scale. Finally, the consumers had to provide their perception of the salt level, the fat level, the tenderness, and the color intensity of the product using 5-points Just-About-Right (JAR) scales.

After the sensory evaluation of the product, the consumers answered a few questions concerning their motivations for having purchased this product.

2.3.3. Sensory characterization of the ideal product

When the consumers decided to stop purchasing and evaluating products, they connected to TimeSens© and selected the corresponding option. This led them to answer a final questionnaire. In this questionnaire, they had to describe their ideal product using FC descriptions according to the same three sensory modalities used to describe the actual products. For each sensory modality, the following instructions were given to the consumers:

- Visual aspect: “Please describe the visual aspect of an ideal ham in your opinion”
- Texture in mouth: “Please describe the texture in mouth of an ideal ham in your opinion”
- Flavor: “Please describe the taste of an ideal ham in your opinion”.

Some consumers did not answer the final questionnaire, resulting in a final number of 415 evaluations for the ideal product.

2.4. Data analyses

All FC data treatments and analyses were performed using R 4.0.2 (R Core Team, 2020). The lexicon provided with IRaMuTeQ© (Ratinaud, 2014) software was used for lemmatization and part-of-speech tagging.

Since the focus is on IFC, JAR scales and motivations for purchasing the products were not analyzed in this paper.

2.4.1. FC data treatment

2.4.1.1. FC descriptions of the actual products

As FC descriptions were collected in French, all subsequent treatments were performed in French. The descriptors resulting from the treatments were then translated into English for the present paper. The English-French correspondence of the descriptors can be found in the appendix.

The FC datasets from each of the three sensory modalities (visual aspect, texture in mouth, and flavor) were treated separately with the method described in (Mahieu, Visalli, Thomas, & Schlich, 2020) and summarized thereafter. The FC descriptions of the ideal product were not involved in this process.

The descriptions were first cleaned, lemmatized, and filtered. Then, the descriptors with similar meanings were grouped into latent-descriptors relying on an ascendant hierarchical classification.

Among all the descriptors and latent-descriptors, only those mentioned throughout at least 5% of the evaluations of at least one product were retained for further analysis.

Finally, the descriptors were cross-tabulated with the consumers and the products indicating whether each descriptor was cited in the corresponding evaluation or not.

2.4.1.2. FC descriptions of the ideal product

The FC descriptions of the ideal product were treated the same manner as the FC descriptions of the actual products. They were cleaned, lemmatized, and filtered using the same filters that those used for the actual products, and the same descriptor groupings were applied. Some additional descriptors not mentioned for the actual products appeared in the descriptions of the ideal product. However, these additional descriptors were not mentioned by at least 5% of the consumers that described their ideal product and they were thus not retained for further analyses.

Finally, the descriptors were cross-tabulated with the consumers indicating whether each descriptor was cited by the corresponding consumer in its description of the ideal product or not.

2.4.2. Panel level

2.4.2.1. Drivers of liking

The liking scores were regressed against the consumer factor, the product factor, and the descriptor factors using a mixed linear model fitted on all evaluations. Each descriptor factor had two levels: absence or presence, the absence level being the reference one. The descriptor factors and the product factor were considered as fixed while the consumer factor was considered as random. The regression loading of each descriptor was considered as an estimate of its impact on liking scores. Confidence intervals (α = 5%) for the regression loadings were computed using the Satterthwaite approximation (Giesbrecht & Burns, 1985; Hrong-Tai Fai & Cornelius, 1996; Satterthwaite, 1946).

2.4.2.2. Ideal product

The proportion of citations of each mentioned descriptor in the FC descriptions of the ideal product were computed. Confidence intervals (α = 5%) for these proportions were computed based on bootstrap resamplings of the consumers (1000 simulations). Descriptors significantly more frequently cited for the ideal product relatively to the actual products were investigated using multiple-response hypergeometric tests (Mahieu, Schlich, Visalli, & Cardot, 2021) with a one-sided greater alternative hypothesis (α = 5%). For these tests, the random hypergeometric samplings to estimate the null distribution were performed using the FC descriptions of the actual products provided by the consumers having described their ideal product.

For each of the three sensory modalities, a multiple-response Correspondence Analysis (MR-CA) (Mahieu et al., 2021) was performed based on the descriptor citation proportions for the actual products. The ideal product was projected as a supplementary observation (based on its own descriptor citation proportions) into the sensory space depicted by the actual products. Confidence ellipse (α = 5%) for the ideal product coordinates was build based on bootstrap resampling of the consumers

(1000 simulations). Finally, the vector of mean liking scores of the actual products was projected as a supplementary variable into the sensory space by computing its weighted correlation coefficient with the MR-CA axes and using the same weight as the MR-CA. This was performed to link the mean liking scores to the position of the ideal product.

2.4.3. Consumer segments

2.4.3.1. Segmentation of the consumers based on their ideal product

The consumers were segmented based on their FC descriptions of the ideal product considering the three sensory modalities and using a mixture-model-based clustering for nominal data (Linzer & Lewis, 2011). The model assumes the data coming from a finite mixture of K class-conditional probability distributions. The mixing proportions and the class-conditional probability distributions are estimated by maximizing the log-likelihood of the model using the expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). The models ranging from K = 1 class to K = 10 classes were built. The “best” model was selected as the one having the lowest mean of its AIC (Akaike, 1974) and BIC (Schwarz, 1978). This resulted in retaining the two-class model. Finally, each consumer was affected to a class using a *maximum a posteriori* (MAP) decision rule. This resulted in two segments respectively composed of 351 (G1) and 64 (G2) consumers.

2.4.3.2. Characterization of each segment of consumers

Potential differences between the two segments in terms of gender repartition and age group repartition were investigated using a chi-square test (α = 5%). Potential differences between the two segments in terms of average frequency of consumption of cooked hams by month were investigated using a bilateral t-test (α = 5%).

2.4.3.3. Ideal product of each segment of consumers

The same computations as presented in section 2.5.1.2 were performed within each segment.

2.4.3.4. Drivers of liking of each segment of consumers

The same computations as presented in section 2.5.1.1 were performed within each segment. The drivers of liking of each segment were investigated to be compared to the ideal product of the corresponding segment.

3. Results

3.1. Panel level

3.1.1. Drivers of liking

Fig. 1 shows that the identified drivers of liking make sense from a sensory point of view. The negatively connoted descriptors (e.g. *F_insipid*, *T_elastic_rubbery*, etc.) were diagnosed as negative drivers of liking. On the contrary, the positively connoted descriptors (e.g *F_fragrant*, *T_soft_tender_melting*, etc.) were diagnosed as positive drivers of liking. Some less trivial information is also shown in Fig. 1. For example, observing *F_not_salty* as a driver of liking and *F_salty* as a driver of disliking can be useful information, especially in a nutritional context. It appears in Fig. 1 that flavor impacted more liking than the texture in mouth which itself impacted more liking than the visual aspect. Finally, it can be seen in Fig. 1 that there were more drivers of disliking than drivers of liking. Also, drivers of disliking had more impact on liking scores in absolute value than drivers of liking.

3.1.2. Ideal product

Fig. 2 shows that the mentioned descriptors in the FC descriptions of the ideal product were relevant as no negatively connoted descriptors were mentioned, which confirms that the consumers understood the concept of describing their ideal product. Some characteristics appeared very important to be found in the ideal product: *V_not_fat*, *V_pink*, *V_soft_tender*, *F_not_salty*. The descriptors significantly more frequently cited for the ideal product relatively to the actual products were consistent with the identified drivers of liking. However, some differences can still be noticed. One descriptor significantly associated with the ideal product was not identified as a driver of liking: a “natural” visual appearance (*V_natural*). On the contrary, *F_fragrant* identified as a driver of liking was not cited in the FC descriptions of the ideal product. Finally, some opposite descriptors (e.g. *F_salty* vs. *F_not_salty*) were mentioned in

FC descriptions of the ideal product which justifies investigating if consumer segments exist (see Section 3.2).

For the three sensory modalities, Fig. 3 (note that *IdealG1*, *IdealG2*, *LikingG1* and *LikingG2* refer to a subsequent segmentation discussed later in section 3.2) shows that the ideal product achieved the most extreme coordinates in the direction of the liking among all the products and lied in a region of the sensory space that none of the actual products reached. The first point confirms that the consumers understood the concept of describing their ideal product and suggests that they provided ideal product descriptions consistent with their liking scores. The second point suggests that none of the actual products was ideal and that gathering descriptions of the ideal product can provide relevant information. It is worth noticing that even if the ideal product lied in a particular region of the sensory space, it was not the most distant product from the average. This statement is true for the three sensory modalities and suggests that the ideal product might be realistic. Interestingly, the confidence ellipse of the ideal product was larger for the flavor modality than for the two other sensory modalities. This is likely because consumers were more consensual in describing their ideal product regarding visual aspect and texture in mouth than regarding flavor and reinforces that investigating if consumer segments exist might be relevant.

Fig. 3 also suggests that the flavor modality is the most important regarding hedonic appreciation. This is further confirmed by the average absolute weighted correlation of the mean liking scores with the whole sensory axes: 0.139 for visual aspect, 0.201 for texture in mouth, and 0.261 for flavor. Finally, it is interesting to notice that this ranking of the sensory modalities regarding the link between their sensory axes and the mean liking scores is the same as the ranking observed for the drivers of liking regarding the impact of each sensory modality on the liking scores.

3.2. Consumer segments

3.2.1. Characterization of each segment of consumers

The two segment were not statistically different regarding their gender repartition ($Chi^2 = 0.074$, $df = 1$, $p = 0.7857$), their age group repartition ($Chi^2 = 2.771$, $df = 2$, $p = 0.2502$), and their average frequency of consumption of cooked hams by month ($t = -0.5802$, $df = 374$, $p = 0.5621$).

3.2.2. Ideal product of each segment of consumers

Fig. 4 shows that the two segments of consumers are interpretable. The ideal products of the two segments mainly differed regarding their flavor. The ideal product of G1 was described as *F_not_salty* approximately half of the time while it was never described as *F_salty*. On the contrary, the ideal product of G2 was always described as *F_salty* while it was never described as *F_not_salty*. This suggests that two types of consumers exist. Those that would like their ideal product not to be salty and those that would like their ideal product to be salty, the “salty lovers” being fewer (≈ 15% of the consumers) than the others. Other smaller differences can be noticed between the ideal products of the two segments: the ideal product of G1 was more often described as *F_ham_taste* and *F_spicy_stocks_aromatics* than the one of G2.

Fig. 3 confirms the results from Fig. 4: the ideal products of the two segments differed regarding their flavor but neither their texture in mouth nor their visual aspect. Not surprisingly, regarding the flavor modality, the two ideal products were opposed on the second dimension, which was a gradient of saltiness. Fig. 3 shows that the ideal product of G1 is very close and thus similar to that of the panel. This makes sense since G1 represents an overwhelming majority as compared to G2. Regarding, the mean liking scores, the two segments appeared to have a similar pattern, close to that of the panel. G1 seemed more consistent than G2 because its ideal product is located farther away in the direction of its mean liking scores for the flavor modality.

3.2.3. Drivers of liking of each segment of consumers

Fig. 5 shows that the drivers of liking of each segment were only partially consistent with their corresponding ideal product. Regarding saltiness, which was the main difference between the two ideal products, the drivers of liking of G1 were perfectly consistent with its ideal product: *F_not_salty* was a driver of liking and *F_salty* was a driver of disliking. For G2, the loading of *F_salty* was positive as opposed to this same loading for G1, but not significant. The loading of *F_not_salty* was also not significant but it was positive and higher than that of *F_salty*. This reinforces the evoked doubt (Fig. 3) on the consistency of G2. Regarding *F_ham_taste* and *F_spicy_stocks_aromatics*, which were the other main differences between the two ideal products, *F_ham_taste* was a driver of liking for the two segments and with the same intensity and *F_spicy_stocks_aromatics* was a driver of liking for G1 and not

G2. However, this difference between G1 and G2 might be due to the different number of consumers in the two segments which led to the confidence intervals of G2 being larger than for G1. The fact that the loading of *F_spicy_stocks_aromatics* was higher for G2 than for G1 reinforces this line of reasoning. Overall, the main differences between the two ideal products were only moderately recovered by comparing the drivers of liking of each segment. However, regarding the most important difference, which was the level of saltiness, G1 had drivers of liking consistent with its ideal product and a trend of consistency existed for G2 since its loading for *F_salty* was positive as opposed to G1.

4. Discussion

4.1. Drivers of liking vs. ideal product

The ideal product and the drivers of liking are different approaches that have their benefits and drawbacks. The drivers of liking are implicit and thus not subject to cognitive and attitudinal bias unlike ideal product descriptions (Li, Hayes, & Ziegler, 2015). However, drivers of liking depend on the actual product space. This constraint could result in some loss and/or some misleading information if too many sensory characteristics are confused and/or not well represented by the actual product space. Since the ideal product does not depend directly on the actual product space, it enables exploring a larger sensory space than that depicted by the actual products (Worch, Crine, Gruel, & Lê, 2014).

Overall, the ideal product and the drivers of liking should be considered complementary rather than competitors: they reinforce and validate each other. Drivers of liking which are significantly and frequently associated with the ideal product are definitely important characteristics regarding appreciation. In the specific context of FC, they are even more complimentary since some obvious and logical characteristics (e.g. *F_fragrant* in this study) may not be mentioned in the descriptions of the ideal product, as they are essential and natural. On the contrary, some characteristics confused and/or rarely present in the actual products (e.g. *V_natural* in this study) can be caught only thanks to the ideal product characterization.

In this study, drivers of liking and the panel's average ideal product provided information in agreement with each other. This suggests that this information can be used from a product development point of view. Especially, including less salt in the manufacturing process of the cooked hams would be beneficial from a nutritional point of view and could possibly increase hedonic appreciation, but certainly not decrease it.

4.2. Panel level vs. consumer segments for the ideal product

To the best of our knowledge, only one study previously proposed to segment the consumers based on their ideal product (Chan, Kwong, & Hu, 2012). Segmenting the consumers based on their ideal product makes sense only in two situations. The first one is when opposite descriptors (e.g. *salty* vs. *not_salty*) are used in individual ideal product descriptions. The second case is when the description of the ideal product is highly variable among consumers. To determine if segmenting the consumers is relevant, and when it is, the number of segments to consider should be determined using objective criterions. Depending on the strategy of clustering adopted, different criterions exist. When mixture models are used, as in this study, information criterions such as AIC (Akaike, 1974) and BIC (Schwarz, 1978) can be used. When hierarchical clustering and/or *k*-means algorithm are used, quality of clustering indexes such as the Silhouette index (Rousseeuw, 1987) and the Gap statistic (Tibshirani, Walther, & Hastie, 2001) can be used.

Even when segmenting the consumers based on their ideal product appears relevant from both a qualitative and a statistical point of view, checking the consistency of each segment is important (Brard & Lê, 2016; Worch et al., 2014; Worch, Lê, Punter, & Pagès, 2012a, 2012b). If the ideal product of one or more segments does not make sense regarding their drivers of liking, segmenting the consumers is questionable. Similarly, when the segments share common drivers of liking but have a different ideal product, segmenting is questionable. In this context, to better understand the differences between the ideal products of each segment, using mapping techniques (e.g. factorial analyses) and absolute measurements (e.g. probabilities of citations) are useful and should be used conjointly. Further, considering that some consumers could eventually provide ideal product descriptions based on non-sensory criteria (e.g. health) (Worch et al., 2013) could help understanding some non-consistent segments. Indeed some consumers could like sweet products but their ideal product could be described as not sweet because they are diabetics for example. However, since the ideal descriptions are instructed to be provided based on the sensory perception (visual aspect, texture in mouth and flavor in this study) this is unlikely to occur.

If different segments of consumers are identified, but one or some of them are of a too-small size, then one should not consider the segmentation (Worch et al., 2012a, 2012b).

In the present paper, G1 highly dominated G2 in terms of size. Further, the consistency of G2 was highly questionable, and G1 and G2 had no clear difference in their drivers of liking except maybe on the level of saltiness. This suggests that for this paper, the analyses performed at the panel level considering a single ideal product are likely the most relevant. Alternatively, as suggested by (Worch et al., 2012a, 2012b), the ideal product descriptions coming from the consumers of G2 could be dropped from the analysis by considering only those from G1.

4.3. Limitations

A first limitation comes from the uncommon data collection procedure of this study. Indeed, to the best of our knowledge, it is the first time that sensory and hedonic data are gathered from consumers purchasing the products they evaluate, which resulted in unbalanced data for the actual products. This uncommon procedure does not appear to be a major limitation as the data make sense. However, it worth emphasizing that the liking scores of the actual products may have been overestimated. Indeed, because consumers selected the products they evaluated, some of them may have selected products they usually purchase and like. Knowing that 20% of the evaluations among the 2758 ones were performed on usually purchased hams and that an average overall liking score of 6.35 (all products combined) was observed, the previous assertion could be at least partly verified. However, other strategies of selection from the consumers may have occurred such as selecting less expensive ones to maximize income from compensations, testing more expensive ones as they were partly refunded by the compensations, or selecting hams based on their labels and/or allegations. These other strategies, considered together with the requirement that, for being compensated, the consumers had to evaluate at least 4 different hams from the list, are the most likely explanations to the fact that most of the hams belonging to the list were evaluated a fair number of times, thus limiting the liking overestimation. Another point worth emphasizing is that, since the consumers selected their evaluated products, they may have restricted the product space and with that, the range of encountered sensory characteristics, which may have affected the ideal product descriptions provided after the evaluations of actual products. Indeed, consumers likely defined what they like and dislike based on the evaluations of actual products. Depending on the practitioners' aims, if gathering less "informed" ideal product descriptions is of interest, consumers could be instructed to provide them before evaluations of actual products but this could inversely affect actual products descriptions. Anyway, investigating the method presented in this paper with a more "conventional" experimental procedure might be an interesting direction for some future research. In particular, comparing the consumer segments resulting from a segmentation on either ideal product data or liking data would be of great interest. Segmenting consumers based on liking data was not performed in this study because of the uncommon experimental design that resulted in a "product by consumer" matrix of liking scores having 81% of missing data.

A second limitation comes from the IFC method and the data analysis procedure proposed in this study. More specifically, if some descriptors not present in the FC descriptions of the actual products are mentioned in the FC descriptions of the ideal product, the projection of the ideal product into the sensory space depicted by the actual products can only be performed on basis of the descriptors shared by the actual and the ideal product descriptions. However, this is not a major limitation since the aim of this projection is to investigate the position of the ideal product relative to the actual products, which make sense to be performed on the same set of descriptors. All the other analyses presented in this study can be performed equivalently with additional descriptors for the ideal product as compared to the actual products. Finally, it has to be mentioned that if this situation occurs, it is a nice argument in favor of IFC since no other existing method can investigate the hedonic importance of descriptors not present within the actual product space.

5. Conclusion

The paper proposes to use Free-Comment (FC) sensory data to be used in the well-established link between sensory and hedonic data. Further, it introduced a new methodology called Ideal-Free-Comment (IFC) where consumers are instructed to describe actual products and then their ideal product thanks to FC. This enables investigating drivers of liking and characterizing the ideal product without the use of a pre-established list of descriptors, which *de facto* avoids inherent limitations to any pre-established list. Further, since the characterization of the ideal product is directly performed, it does not depend on the actual product space, and the hedonic importance of descriptors confused and/or rarely present in the actual products can thus be investigated. Identification of drivers of liking based on FC data and IFC were used on cooked hams with consumers purchasing the products they evaluated at home and it showed relevant results. Drivers of liking based on FC data and IFC provide sensory analysts with new complementary tools to understand consumers' hedonic appreciation without the use of a pre-established list of descriptors.

Acknowledgments

This study is part of a Ph.D. financed by the Region Bourgogne-Franche-Comté and the SensoStat Company.

References

Adams, J., Williams, A., Lancaster, B., & Foley, M. (2007). Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. In, *7th Pangborn Sensory Science Symposium*. Minneapolis, USA.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723.

Ares, G., Dauber, C., Fernández, E., Giménez, A., & Varela, P. (2014). Penalty analysis based on CATA questions to identify drivers of liking and directions for product reformulation. *Food Quality and Preference, 32*, 65-76.

Ares, G., de Andrade, J. C., Antúnez, L., Alcaire, F., Swaney-Stueve, M., Gordon, S., et al. (2017). Hedonic product optimisation: CATA questions as alternatives to JAR scales. *Food Quality and Preference, 55*, 67-78.

Ares, G., Jaeger, S. R., Bava, C. M., Chheang, S. L., Jin, D., Gimenez, A., et al. (2013). CATA questions for sensory product characterization: Raising awareness of biases. *Food Quality and Preference, 30*(2), 114-127.

Ares, G., & Varela, P. (2017). Trained vs. consumer panels for analytical testing: Fueling a long lasting debate in the field. *Food Quality and Preference, 61*, 79-86.

Ares, G., Varela, P., Rado, G., & Giménez, A. (2011). Identifying ideal products using three different consumer profiling methodologies. Comparison with external preference mapping. *Food Quality and Preference, 22*(6), 581-591.

Brard, M., & Lê, S. (2016). The Ideal Pair Method, an Alternative to the Ideal Profile Method Based on Pairwise Comparisons: Application to a Panel of Children. *Journal of Sensory Studies, 31*(4), 306-313.

Bruzzone, F., Vidal, L., Antúnez, L., Giménez, A., Deliza, R., & Ares, G. (2015). Comparison of intensity scales and CATA questions in new product development: Sensory characterisation and directions for product reformulation of milk desserts. *Food Quality and Preference, 44*, 183-193.

Callegaro, M., Murakami, M. H., Tepman, Z., & Henderson, V. (2015). Yes-no answers versus check-all in self-administered modes. *International Journal of Market Research, 57*, 203-223.

Carroll, J. D. (1972). Individual Differences and Multidimensional Scaling. In, *R. N. Shepard, A. K. Romney, & Si Nerloves (Eds.), Multidimensional scaling: Theory and applications in the behavioral sciences*. New York: Academic Press.

Chan, K. Y., Kwong, C. K., & Hu, B. Q. (2012). Market segmentation and ideal point identification for new product design using fuzzy data compression and fuzzy clustering methods. *Applied Soft Computing, 12*(4), 1371-1378.

Coulon-Leroy, C., Symoneaux, R., Lawrence, G., Mehinagic, E., & Maitre, I. (2017). Mixed Profiling: A new tool of sensory analysis in a professional context. Application to wines. *Food Quality and Preference, 57*, 8-16.

Danzart, M. (2009). *SSHA 3eme (Ed.), Evaluation sensorielle. Manuel méthodologique*. Paris: Lavoisier.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1-38.

Dooley, L., Lee, Y.-s., & Meullenet, J.-F. (2010). The application of check-all-that-apply (CATA) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. *Food Quality and Preference, 21*(4), 394-401.

Escofier, B., & Pagès, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis, 18*(1), 121-140.

Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A., & Nicod, H. (2006). An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality and Preference, 17*(7), 604-614.

Giesbrecht, F. G., & Burns, J. C. (1985). Two-Stage Analysis Based on a Mixed Model: Large-Sample Asymptotic Theory and Small-Sample Simulation Results. *Biometrics, 41*(2), 477-486.

Greenhoff, K., & MacFie, H. J. H. (1994). Preference mapping in practice. In, *H. J. H. MacFie & D. M. H. Thompson (Eds.), Measurement of food preferences*. London: Blackie Academic & Professionals.

Grygorczyk, A., Lesschaeve, I., Corredig, M., & Duizer, L. (2013). Extraction of consumer texture preferences for yogurt: Comparison of the preferred attribute elicitation method to conventional profiling. *Food Quality and Preference, 27*(2), 215-222.

Hrong-Tai Fai, A., & Cornelius, P. L. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced

split-plot experiments. *Journal of Statistical Computation and Simulation, 54*(4), 363-378.
Kim, I.-A., Hopkinson, A., van Hout, D., & Lee, H.-S. (2017). A novel two-step rating-based 'double-faced applicability' test. Part 1: Its performance in sample discrimination in comparison to simple one-step applicability rating. *Food Quality and Preference, 56*, 189-200.
Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.
Lagrange, V., & Norback, J. P. (1987). Product optimization and the Acceptor Set Size. *Journal of Sensory Studies, 2*(2), 119-136.
Lawless, H. T., & Heymann, H. (1999). *Sensory evaluation of food: Principles and practices*. New York: Kluwer Academic/Plenum Publishers.
Li, B., Hayes, J. E., & Ziegler, G. R. (2015). Maximizing overall liking results in a superior product to minimizing deviations from ideal ratings: An optimization case study with coffee-flavored milk. *Food Quality and Preference, 42*, 27-36.
Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software, 42*(10).
Luc, A., Lê, S., & Philippe, M. (2020). Nudging consumers for relevant data using Free JAR profiling: An application to product development. *Food Quality and Preference, 79*.
Mahieu, B., Schlich, P., Visalli, M., & Cardot, H. (2021). A multiple-response chi-square framework for the analysis of Free-Comment and Check-All-That-Apply data. *Food Quality and Preference, 93*.
Mahieu, B., Visalli, M., Thomas, A., & Schlich, P. (2020). Free-comment outperformed check-all-that-apply in the sensory characterisation of wines with consumers at home. *Food Quality and Preference, 84*.
McEwan, J. A. (1996). Preference Mapping for product optimization. In, *T. Naes & E. Risvik (Eds.), Multivariate analysis of data in sensory science*. New York: Elsevier.
Meilgaard, M., Civille, G. V., & Carr, B. T. (1991). *Sensory Evaluation Techniques*. Boca Raton, FL: CRC Press.
Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference, 30*(2), 309-319.
Moskowitz, H. R. (1972). Subjective ideals and sensory optimization in evaluating perceptual dimensions in food. *Journal of Applied Psychology, 56*(1), 60-66.
Popper, R. (2014). Use of Just-About-Right Scales in Consumer Research. In, *P. Varela & G. Ares (Eds), Novel techniques in sensory characterization and consumer profiling*. Boca Raton, FL (United States): CRC Press.
R Core Team. (2020). R: A language and environment for statistical computing. In. Vienna, Austria: R Foundation for Statistical Computing.
Ratinaud, P. (2014). IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. In. France.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65.
Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin, 2*(6), 110-114.
Schlich, P., & McEwan, J. A. (1992). Cartographie des préférences. Un outil statistique pour l'industrie agro-alimentaire. *Science des Aliments, 12*, 339-355.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.
Stone, H., & Sidel, J. L. (1993). *Sensory evaluation practices*. California: Academic Press.
ten Kleij, F., & Musters, P. A. D. (2003). Text analysis of open-ended survey responses: a complementary method to preference mapping. *Food Quality and Preference, 14*(1), 43-52.
Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63*(2), 411-423.
Valentin, D., Chollet, S., Lelièvre, M., & Abdi, H. (2012). Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology, 47*(8), 1563-1578.
van Kleef, E., van Trijp, H. C. M., & Luning, P. (2006). Internal versus external preference analysis: An exploratory study on end-user evaluation. *Food Quality and Preference, 17*(5), 387-399.
van Trijp, H. C. M., Punter, P. H., Mickartz, F., & Kruithof, L. (2007). The quest for the ideal product: Comparing different methods and approaches. *Food Quality and Preference, 18*(5), 729-740.
Varela, P., Antúnez, L., Carlehög, M., Alcaire, F., Castura, J. C., Berget, I., et al. (2018). What is dominance? An exploration of the concept in TDS tests with trained assessors and consumers. *Food Quality and Preference, 64*, 72-81.
Varela, P., & Ares, G. (2012). Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International, 48*(2), 893-908.
Worch, T. (2013). PrefMFA, a solution taking the best of both internal and external preference mapping techniques. *Food Quality and Preference, 30*(2), 180-191.
Worch, T., Crine, A., Gruel, A., & Lê, S. (2014). Analysis and validation of the Ideal Profile Method: Application to a skin cream study. *Food Quality and Preference, 32*, 132-144.
Worch, T., Lê, S., Punter, P., & Pagès, J. (2012a). Assessment of the consistency of ideal profiles according to non-ideal data for IPM. *Food Quality and Preference, 24*(1), 99-110.
Worch, T., Lê, S., Punter, P., & Pagès, J. (2012b). Extension of the consistency of the data obtained with the Ideal Profile Method: Would the ideal products be more liked than the tested products? *Food Quality and Preference, 26*(1), 74-80.
Worch, T., Lê, S., Punter, P., & Pagès, J. (2013). Ideal Profile Method (IPM): The ins and outs. *Food Quality and Preference, 28*(1), 45-59.

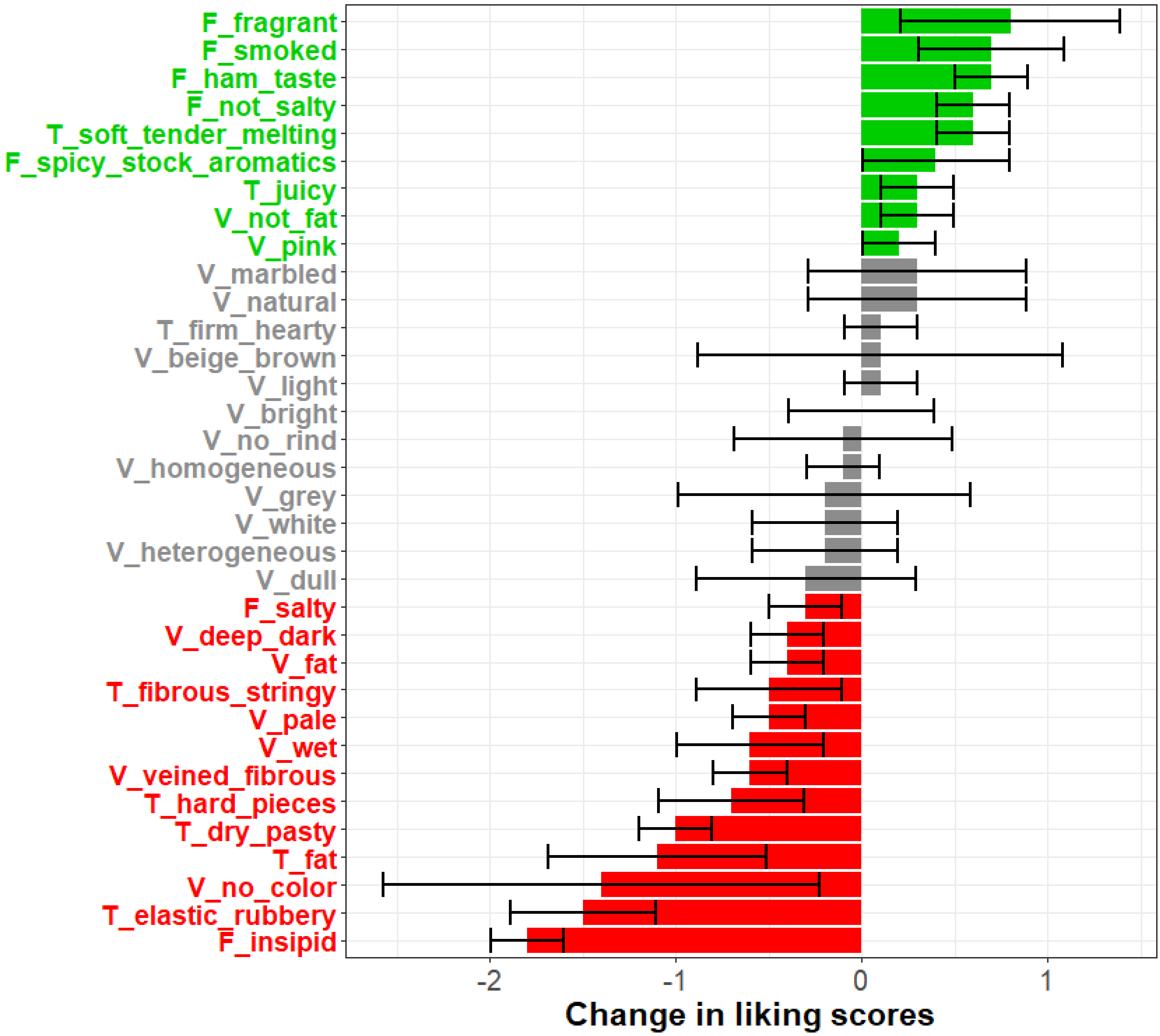
Fig. 1: Regression loadings of each descriptor with their respective confidence intervals (α = 5%). V stands for the visual descriptors, T stands for the texture in mouth descriptors and F stands for the flavor descriptors. Green (resp. red) bars represent significant (α = 5%) positive (resp. negative) drivers of liking.

Fig. 2: Proportions of citation of descriptors mentioned in the FC descriptions of the ideal product with their respective confidence intervals ($\alpha = 5\%$). Green bars represent descriptors significantly more frequently cited for the ideal product relatively to the actual products (multiple-response hypergeometric test, $\alpha = 5\%$).

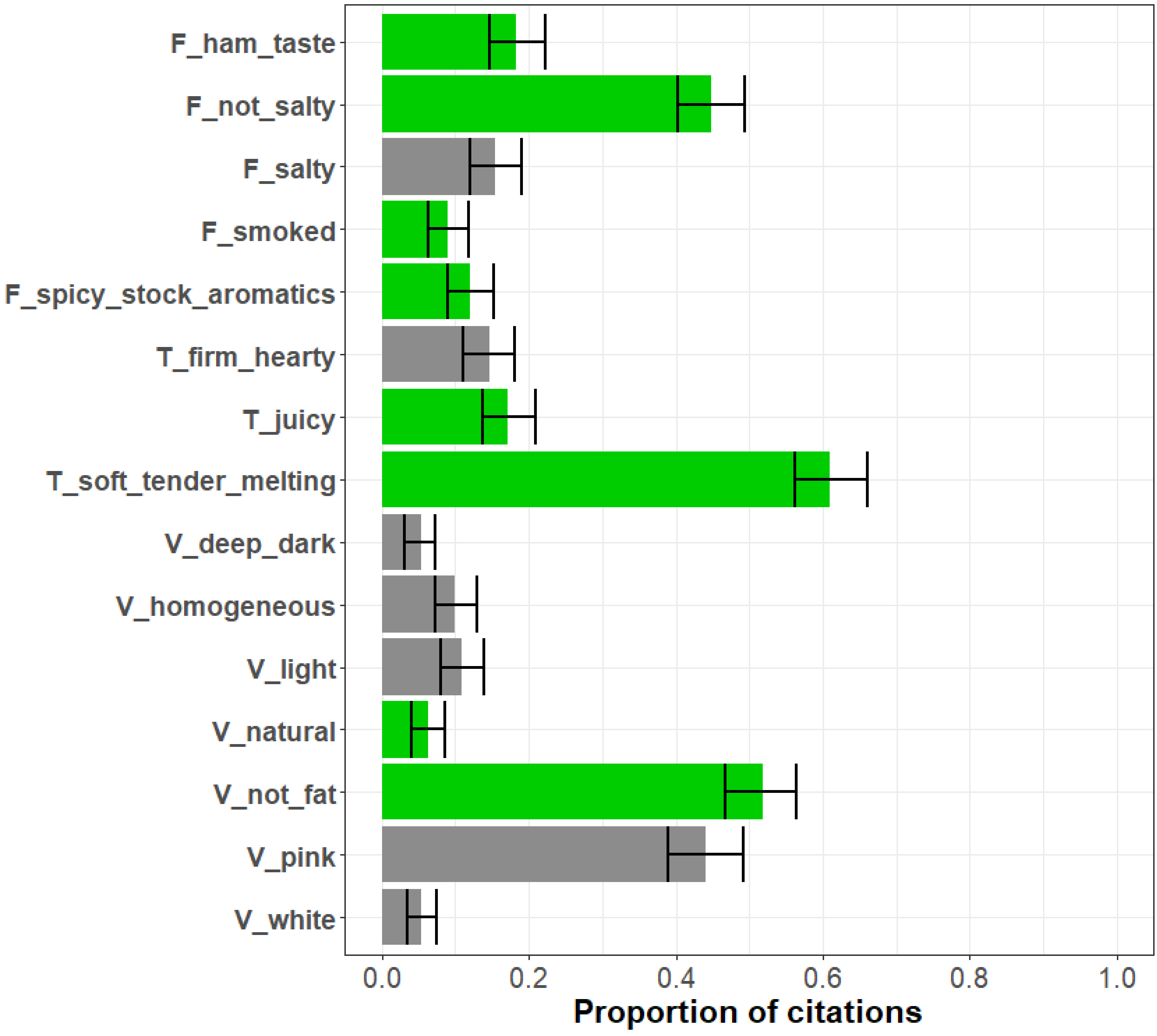
Fig. 3: Biplot from multiple-response Correspondence Analysis with the panel (Ideal) and by segment (IdealG1 and IdealG2) ideal products (projected as supplementary observation), their confidence ellipse (α = 5%) and the mean panel (Liking) and by segment (LikingG1 and LikingG2) liking scores (projected as supplementary variable): (a) axes 1-2 visual aspect, (b) axes 3-2 visual aspect, (c) axes 1-2 texture in mouth, (d) axes 3-2 texture in mouth, (e) axes 1-2 flavor, (f) axes 3-2 flavor. Blue points are the actual products (unlabeled for sake of readability). Weighted correlation values of liking scores can be read thanks to the axes ticks.

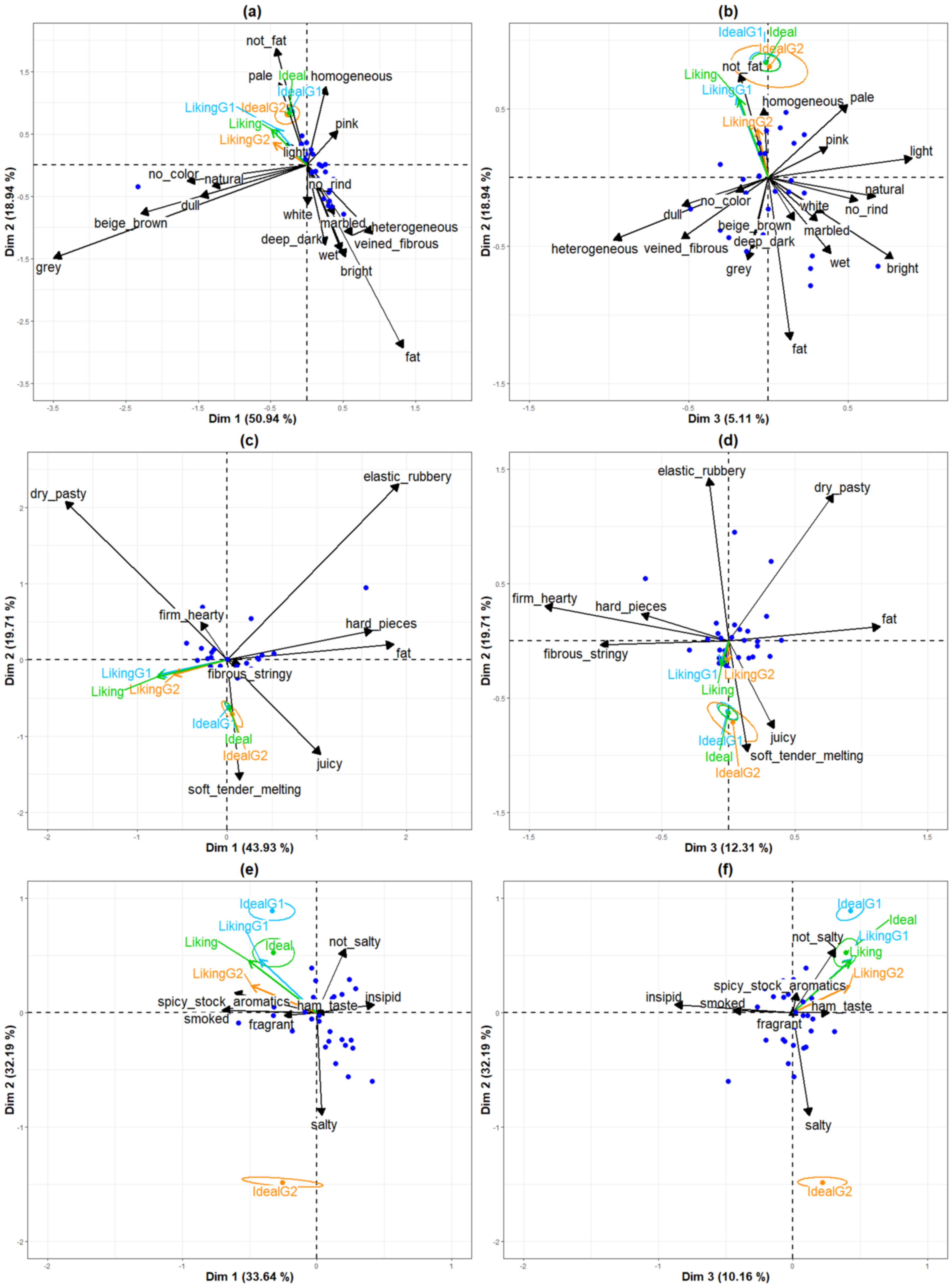
Fig. 4: Proportion of citations of descriptors mentioned in the FC descriptions of the ideal product within each segment with their respective confidence intervals (α = 5%).

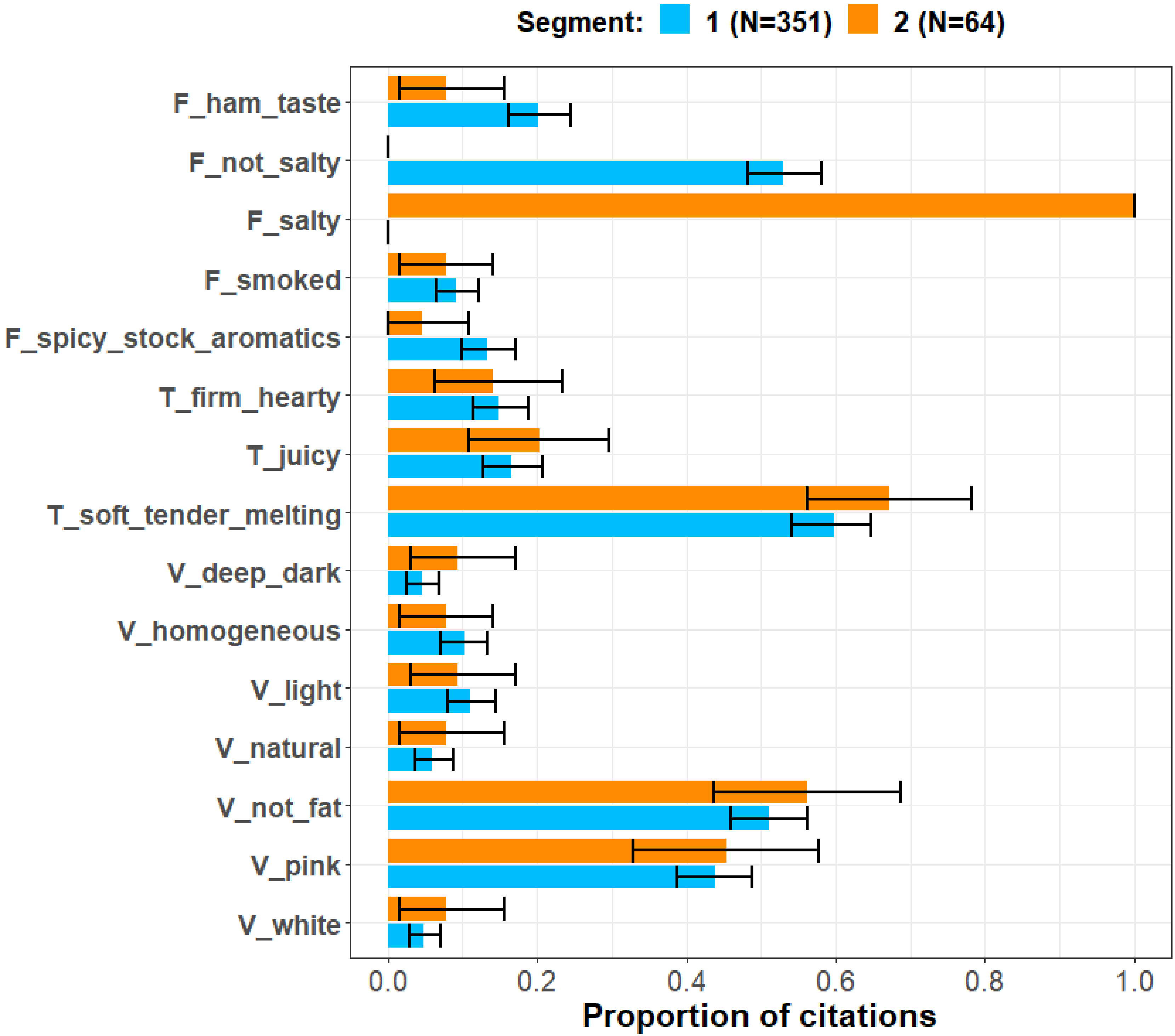
Fig. 5: Regression loadings of each descriptor with their respective confidence intervals (α = 5%) for the two segments of consumers: (a) G1 (N = 351) and (b) G2 (N = 64). V stands for the visual descriptors, T stands for the texture in mouth descriptors and F stands for the flavor descriptors. Green (resp. red) bars represent significant (α = 5%) positive (resp. negative) drivers of liking.

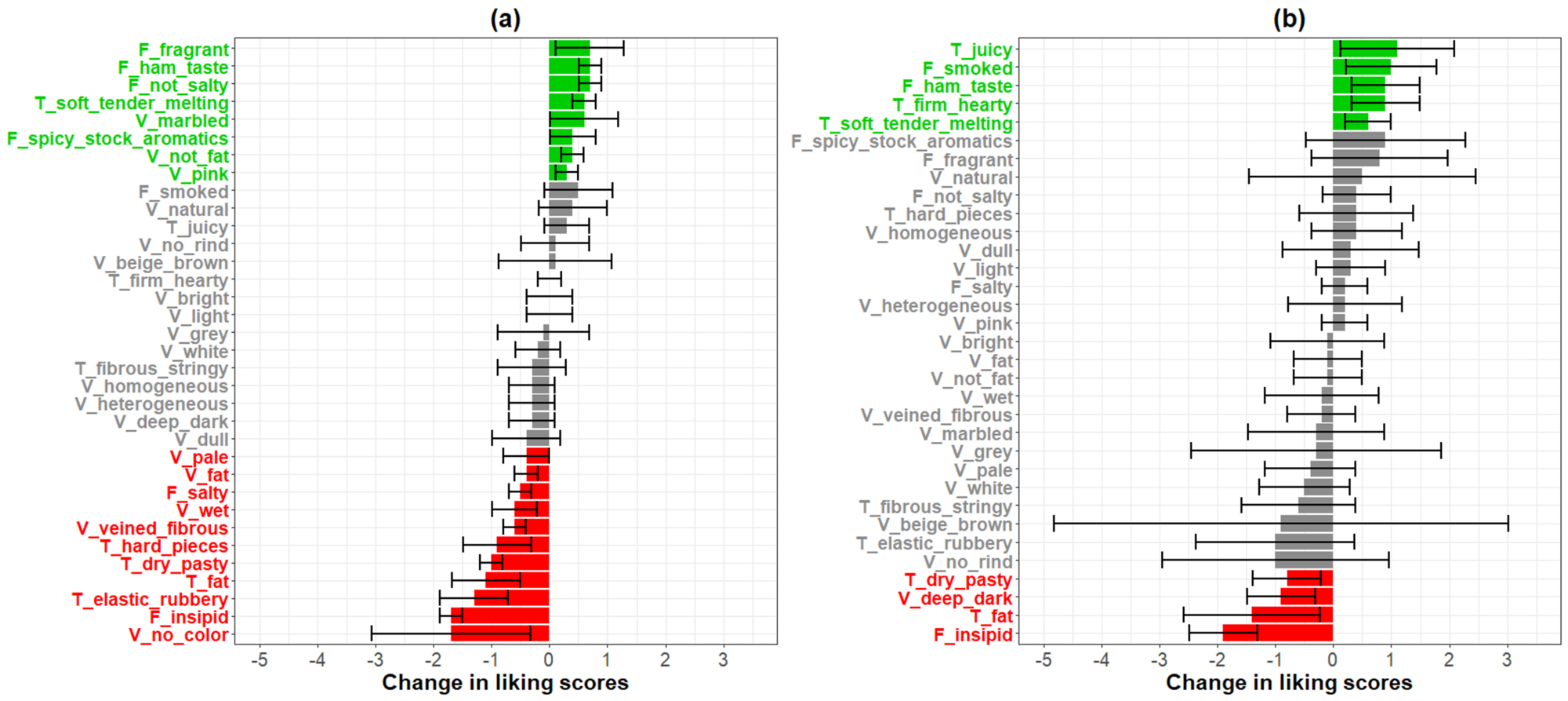


F_fragrant
F_smoked
F_ham_taste
F_not_salty
T_soft_tender_melting
F_spicy_stock_aromatics
T_juicy
V_not_fat
V_pink
V_marbled
V_natural
T_firm_hearty
V_beige_brown
V_light
V_bright
V_no_rind
V_homogeneous
V_grey
V_white
V_heterogeneous
V_dull
F_salty
V_deep_dark
V_fat
T_fibrous_stringy
V_pale
V_wet
V_veined_fibrous
T_hard_pieces
T_dry_pasty
T_fat
V_no_color
T_elastic_rubbery
F_insipid
-2
-1
0
1
Change in liking scores


F_ham_taste
F_not_salty
F_salty
F_smoked
F_spicy_stock_aromatics
T_firm_hearty
T_juicy
T_soft_tender_melting
V_deep_dark
V_homogeneous
V_light
V_natural
V_not_fat
V_pink
V_white
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of citations


(a)
(b)
(c)
(d)
(e)
(f)
Dim 1 (50.94 %)
Dim 2 (18.94 %)
Dim 3 (5.11 %)
Dim 1 (43.93 %)
Dim 2 (19.71 %)
Dim 3 (12.31 %)
Dim 1 (33.64 %)
Dim 2 (32.19 %)
Dim 3 (10.16 %)
not_fat
pale
Ideal
homogeneous
IdealG1
LikingG1
IdealG2
Liking
LikingG2
pink
light
no_color
natural
dull
beige_brown
grey
no_rind
white
marbled
heterogeneous
deep_dark
veined_fibrous
wet
bright
fat
elastic_rubbery
dry_pasty
firm_hearty
hard_pieces
fibrous_stringy
juicy
soft_tender_melting
not_salty
insipid
spicy_stock_aromatics
ham_taste
smoked
fragrant
salty


Segment: 1 (N=351) 2 (N=64)
F_ham_taste
F_not_salty
F_salty
F_smoked
F_spicy_stock_aromatics
T_firm_hearty
T_juicy
T_soft_tender_melting
V_deep_dark
V_homogeneous
V_light
V_natural
V_not_fat
V_pink
V_white
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of citations



(a)
F_fragrant
F_ham_taste
F_not_salty
T_soft_tender_melting
V_marbled
F_spicy_stock_aromatics
V_not_fat
V_pink
F_smoked
V_natural
T_juicy
V_no_rind
V_beige_brown
T_firm_hearty
V_bright
V_light
V_grey
V_white
T_fibrous_stringy
V_homogeneous
V_heterogeneous
V_deep_dark
V_dull
V_pale
V_fat
F_salty
V_wet
V_veined_fibrous
T_hard_pieces
T_dry_pasty
T_fat
T_elastic_rubbery
F_insipid
V_no_color
-5 -4 -3 -2 -1 0 1 2 3
Change in liking scores
(b)
T_juicy
F_smoked
F_ham_taste
T_firm_hearty
T_soft_tender_melting
F_spicy_stock_aromatics
F_fragrant
V_natural
F_not_salty
T_hard_pieces
V_homogeneous
V_dull
V_light
F_salty
V_heterogeneous
V_pink
V_bright
V_fat
V_not_fat
V_wet
V_veined_fibrous
V_marbled
V_grey
V_pale
V_white
T_fibrous_stringy
V_beige_brown
T_elastic_rubbery
V_no_rind
T_dry_pasty
V_deep_dark
T_fat
F_insipid
-5 -4 -3 -2 -1 0 1 2 3
Change in liking scores