Evaluation of omission testing as a method for identifying important odorants in a mixture

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Kirsten Weiss

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Zata Vickers

September, 2016

Acknowledgements

First, I would like to express my deepest thanks to my advisor, Zata Vickers, for her endless help, generous advice, and full support. Your encouragement and expert guidance made it possible for me to complete my graduate degree.

I wish to also thank Gary Reineccius and Peter Sorensen for serving on my examining committee.

I cannot begin to express my thanks to my research assistants, Lauren Wisdorf and Violet Wiebel. I couldn't have done it without your help!

My thanks also goes out to the members of my sensory panel for their patience and full cooperation.

Sincere thank you to Kendra Kauppi for generously accommodating my research.

I would like to thank Jean-Paul Schirle-Keller from the Flavor Center at the University of Minnesota for his help with sample analysis.

I would also like to thank Aaron Rendahl, Amy Ko, Xuan Wang, and Yi Liu for their help with the statistical analyses.

I would like to acknowledge the Sensory Center at the University of Minnesota for funding this research.

Many thanks to my research group for their friendship, encouragement, and moral support.

Finally, I would like to thank my family - to my husband, Jeff, and my children, Nick, Andrew, and Elsa for their unconditional love and support. To my mother, Sharon Schmickle for always believing I could do this. Your support was essential to my success!

"It is very obvious that we have very many different kinds of smells, all the way from the odor of violets and roses up to asafetida. But until you can measure their likenesses and differences you can have no science of odor."

~ ALEXANDER GRAHAM BELL, 1914

Abstract

Because all volatile compounds in food products do not contribute to the perceived aroma, a reliable procedure for determining which compounds do contribute is important. The typical procedure used involves instrumental and sensory methods to isolate, identify, and quantify odorous volatiles followed by recombination and omission testing. The principle idea behind omission testing is quite simple; if a volatile compound contributes to the perceived aroma of the model, people will notice if it is removed. If a compound is removed from a full model and no one can detect that it has been removed, it must not be important. When compounds are combined at concentrations with similar perceived intensities, they tend to fuse or blend creating an aroma with unique characteristics unlike those of the individual compounds. In this situation is omission testing valid? The objective of this study was to determine whether people could learn to discriminate between a five-compound odor mixture with all compounds at the same perceived intensity and the same mixture with any one or two of the five compounds removed.

We selected panelists (n = 18) based on their ability to correctly order the perceived intensities of five concentrations of each of 4 compounds (butyric acid, furaneol, methional, and δ-decalactone). During preliminary test sessions we selected, for each panelist separately, concentrations of each of the four compounds that were equivalent in intensity to 5 ppm acetylpropionyl. We then constructed, for each participant, a mixture of the five compounds that were matched in intensity. Panelists then participated in 20 sessions consisting of a series of A-not-A Tests with corrective feedback. During each of these sessions panelists were presented with ten, complete 5-component mixtures, five, 4-component mixtures (each missing a different component), and five, 3-component mixtures (each missing 2 different components). We used signal detection theory and the discriminability index (d') to evaluate the omission testing data by sessions, panelists, and omitted compounds.

The panelists, as a group, were able to discriminate between the full five-compound mixture and mixtures with any one of the five compounds removed after 60 trials. However, low d' values indicated that the mixtures were extremely difficult to discriminate. We screened out 13 of 31 initial panelists because we could not establish intensity matches between the standard and each of the 4 other compounds. Seven of the 18 who qualified were not able to successfully discriminate between the complete mixture and the n-1 mixture even after 200 trials.

Table of Contents

List of Tables

List of Figure

List of Equations

Introduction

Originally, it was believed that all volatile compounds in foods actively participated in aroma perception (Grosch, 2001). As instrumental methods for analyzing and identifying volatile compounds improved, it became clear that this was not the case, and the search for a reliable procedure for determining which compounds did contribute became important. One commonly used procedure involves instrumental and sensory methods to isolate, identify, and quantify odorous volatiles followed by recombination and omission testing (Dunkel et al., 2014). The principle idea behind omission testing is quite simple: if a volatile compound contributes to the perceived aroma of the model, people will notice its removal. If a compound is removed from a full model and no one can detect its absence, it must not be important. When compounds are combined at concentrations with similar perceived intensities, they tend to fuse or blend creating an aroma with unique characteristics unlike those of the individual compounds (Jinks & Laing, 2001; Laing, Panhuber, Willcox, & Pittman, 1984). This study evaluates the effectiveness of omission testing as a method for determining which compounds contribute to the perceived aroma in a mixture of five compounds combined at similar perceived intensities. This study was conducted in two phases. The purpose of the first phase, odor intensity matching, was to determine the concentrations of five odor compounds that matched in perceived intensity for each panelist. The second phase, omission testing, was conducted to determine whether people could learn to discriminate between the five-component mixture and the same mixture with one or two components removed.

1 Literature Review

1.1 Odor Mixture Perception

Odor mixture perception, a complex and poorly understood process, involves interactions at both the peripheral and central levels of processing. Peripheral processing begins when volatile compounds enter the nose and travel to the olfactory epithelium where they stimulate or inhibit receptors located in Olfactory Receptor Neurons (ORN). Olfactory receptors are not compound specific but can interact with multiple volatiles (Buck, 2000), and volatile compounds are affiliated with more than one receptor type (Buck, 2004; Malnic et al., 1999). However a specific receptor's affinity for the various compatible volatiles can differ (Buck, 2004). When an odor compound binds with its affiliated olfactory receptors, a pattern of receptor activity is sent to the glomeruli regions of the olfactory bulb (Buck, 2004). Each ORN is affiliated with one glomerulus, and each glomerulus region only receives signals from one type of receptors. Because odor molecules trigger more than one ORN - and, therefore more than one glomerulus - a spatial pattern of signals is received in the olfactory bulb. What's more, odor compounds bind with ORN at different rates, lending a time component to the pattern as well. This spatiotemporal pattern is sent via mitral and tufted neurons to the olfactory cortex. Perception of odor quality, a result of the spatiotemporal patterns received from the olfactory bulb, occurs at this cortical level (Wilson & Sullivan, 2011).

In simple odor mixtures spatiotemporal patterns interpreted in the piriform cortex retain enough information about each individual compound to allow for elemental processing, the identification of individual mixture components (Jinks & Laing, 2001; Le

Berre et al., 2008). Laing and Willcox (1983) evaluated the quality and intensity of binary mixtures of odorous compounds. Two compounds at five different concentrations each were combined in pairs and evaluated for intensity and quality of the mixture components. The researchers found that perception of the binary mixture depended on the perceived intensities of the unmixed compounds with the quality of the mixture dominated by the component with the strongest intensity. Mixtures of unequal intensity resulted in domination of quality by the compound at the higher perceived intensity. Mixtures with compounds at similar intensities resulted in detection of both compounds. Similar observations have been made by Atanasova et al. (2005), Laing et al. (1984), Olsson (1994), and Olsson (1998). Elemental processing of simple binary mixtures results in addition (both odors are perceived at similar intensities), suppression (one odor is perceived as stronger than the other), or masking (only one odor is perceived) (Laing et al., 1984). However, this finding depends on the perceived intensities of the unmixed compounds, and similar perceived intensity results in the perception of both compounds.

Our ability to perceive odor mixtures has limits. Humans can identify individual compounds in binary mixtures but not in mixtures containing more than three compounds (Jinks & Laing, 1999; Jinks & Laing, 2001; Laing & Francis, 1989; Laing & Glemarec, 1992; Laing & Jinks, 2001). Performance is not improved by training or experience (Livermore & Laing, 1996), familiarity with the mixture components (Jinks & Laing, 1999), or type of compounds (good blenders/poor blenders) (Livermore & Laing, 1998b). We rarely encounter individual odor compounds or even simple binary mixtures but rather complex mixtures often containing hundreds of volatile compounds. As odor

mixtures become more complex elemental processing gives way to synthetic processing, and the mixtures blend or fuse into unique odor objects. When this happens, it is difficult for humans to determine whether a perceived aroma is the result of one volatile compound or a mixture of many (Laing & Francis, 1989; Laing et al. 1984).

1.1.1 Blending and Fusing in Odor Mixtures

Events occurring at both the peripheral and cognitive levels of perception result in synthetic processing, the perception of complex odor mixtures as individual entities or objects. The spatiotemporal patterns processed at the cortical level are the result of activation and inhibition of receptors at the peripheral levels (Jinks & Laing, 2001). The more odor compounds in a mixture the more complex the mixture becomes. When many volatile compounds are present, competition for receptor sites can occur, and the spatiotemporal patterns of input received in the piriform cortex increase in complexity (Buck, 2013). The more complex the pattern, the more likely the mixture is to be processed in a configurable manner, resulting in the perception of a unique odor identity separate from its individual components. (Jinks & Laing, 2001; Stevenson, 2001). The perceived intensity of the mixture components plays a role in mixture blending as does experience and familiarity.

Several studies of odor perception using mice and rats give us some insight on what may be happening during odor processing. Wilson (2003) found that synthetic processing of binary odor mixtures does not occur until the signals reach the piriform cortex in rats and that this process requires some time exposure to the stimulus. In another study using mice, Zou and Buck (2006) found that binary mixtures of odorous

compounds stimulated portions of the piriform cortex that were not stimulated by the individual components. Barnes and associates (2008) discovered that rats were not able to discriminate between odor mixtures with 10 compounds and the same mixture with one compound removed. However, the rats were able to discriminate between the full mixture and one where the missing compound was replaced by a completely different compound. The researchers concluded that the pattern of receptor activity was completed in the rats' piriform cortex when one compound was removed. However, the new pattern produced by the replacement of one compound was detectable. Similar results were found in a study completed by Lovitz and associates (2012) in which the compound removed from a 10 component mixture was replaced with one of three concentrations of a new compound. The rats were not able to discriminate when one compound was removed. However, their ability to discriminate improved as the concentration of the added compound increased.

Livermore and Laing (1998a) investigated the storage of complex odor mixtures in human memory. They trained 26 panelists to identify eight common odor objects including chocolate, cheese, and honey. All of the odor objects were complex mixtures of many odor compounds. During training, the panelists also rated the intensity of the odor objects. Once the panelists were trained, Livermore and Laing presented the odor objects individually or in groups of two to eight to the panelists who were instructed to identify the odors that were present. The odor objects were combined at similar perceived intensities for each panelist as determined by the intensity ratings done during training. The results were similar to those found for identifying individual compounds in

mixtures discussed above; the trained panelists could not identify more than four odor objects in mixtures. Livermore and Laing concluded that the olfactory system processes complex odor mixtures as "discrete entities."

Complex odor mixtures that blend into odor objects have unique characteristics separate from their individual components. Sinding and associates (2013b) used a sorting task to determine if a six-component odor mixture, smelling like red cordial, smelled anything like its individual components. They instructed 73 novice panelists to sort seven samples "according to odor similarity." The samples included the red cordial blend (in which all of the compounds had been combined at similar perceived intensity) and each of the six individual compounds. The results indicated that the odor mixture had been categorized separately and did not smell like any of its components. The researchers concluded that this was evidence of synthetic odor processing. Similar results were found in a previous study using mixtures smelling like red cordial and sub-mixtures of the same compounds (Sinding et al., 2013b). A study completed by Le Berre and associates (2008) found that binary and ternary mixtures smelled more like pineapple and red cordial than their individual components. Barkat et al. (2012) found similar results in a study on mixtures smelling like pineapple.

Jinks and Laing (2001) investigated changes in the odor quality of four compounds as odor mixtures increased in complexity. In the first part of the study, they began by training 14 panelists to recognize the names of four individual compounds with different characteristic odor qualities. The panelists also rated the intensity of each sample using a line scale. The panelists participated in four testing sessions, one for each

of the four compounds. In each session, the panelists were instructed to smell 15 samples and determine if the compound of interest was present or not. The samples consisted of the four individual compounds as well as all binary, ternary, and quaternary mixtures of the same four compounds at similar perceived intensities (determined during training). The panelists were not able to identify more than two compounds, and, in a mixture of all four, the panelists could not identify any compounds. Jinks and Laing concluded that the spatial patterns used to identify individual compounds, due to increased competition at the receptors as more compounds were added to the mixture, were altered too much to be recognized. Instead, a new pattern for a new odor object was created.

Changes in the perceived intensity of odor mixture components can change the pattern of stimuli received in the piriform cortex resulting in perceptible changes in odor quality (Buck, 2004). In the second part of the study conducted by Jinks and Laing (2001) the panelists evaluated the same 15 samples for odor quality. Each of the 15 samples was evaluated three times for a total of 45 samples. The panelists were given a list of 146 descriptors and were instructed to choose which could be identified in the sample. They were also instructed to rate the intensity of the descriptor using a 9-point scale. The panelists used some, but not all, of the same descriptors they had used for the individual compounds when evaluating the mixtures. No new descriptors were used to evaluate the quality of the mixtures, but the intensity of the qualities changed. Jinks and Laing concluded that both intensity and quality are used to identify odors at the cortical level of processing. A study completed by Le Berre and associates (2008) also found that

small changes in the concentrations of individual mixture components resulted in changes in the odor quality of a ternary mixture smelling like pineapple.

1.1.2 Recognition and Discrimination

Our ability to recognize odors is the result of the processing of the spatiotemporal patterns received in the olfactory piriform cortex (Howard, Plailly, Grueschow, Haynes, & Gottfried, 2009; Stevenson & Wilson, 2007a). Perception in the piriform cortex relies on comparisons of incoming patterns of neural activity with those stored in memory. When an incoming pattern matches one stored in memory, recognition occurs. However, no two odors are ever the same. Subtle changes (alterations in the identity or concentration of mixture components, intensity of inhalation, environmental changes, etc.) result in partial or degraded spatiotemporal patterns received in the piriform cortex. Despite these changes, the olfactory cortex allows for the recognition of odor mixtures. One theory on how this occurs is through pattern completion. In pattern completion, a familiar pattern of incoming stimuli will trigger a strong match in stored memory and the brain will fill in any gaps, completing the pattern and allowing for perceptual stability. (Wilson & Sullivan, 2011)

Our ability to discriminate between odors also relies on comparisons of incoming patterns of neural activity with those stored in memory (Howard et al., 2009; Stevenson & Wilson, 2007a). A familiar pattern of incoming stimuli will trigger a strong match in stored memory inhibiting confusion with other possible matches (Stevenson & Mahmut, 2013). On the other hand, unfamiliar patterns may trigger partial matches with many odor memories, enhancing confusion, and impairing our ability to discriminate

(Stevenson & Boakes, 2003). Furthermore, the task of discriminating between two unfamiliar odors requires that one be stored in short-term memory while the other is evaluated. An unfamiliar odor stored in short term memory is more likely to be altered during the time it takes to complete testing, making discrimination more difficult (Rabin, 1988; Stevenson & Mahmut, 2013). People are not able to discriminate between samples when the odors are new or novel, but they are quite good at discriminating once the odors become familiar (Rabin, 1988).

1.2 Learning and Odor Mixtures

People are better able to discriminate odor quality following training or experience (Jehl, Royet, & Holley, 1995; Rabin, 1988)(Rabin, 1988). Stevenson and Mahmut (2013) conducted a study in which panelists had to describe familiar and unfamiliar odor mixtures. They were asked to match samples with their descriptions immediately after creating them and then again later. There were no differences in matching abilities immediately after describing familiar and unfamiliar mixtures, but, after a few minutes, they could match the familiar odors with descriptions more frequently than unfamiliar odors. Therefore, Stevenson and Mahmut concluded that people are better able to discriminate when we have more experience with an odor.

Perceptual learning for the purpose of identification and discrimination occurs through a process of categorization. Each unfamiliar or novel sample or stimuli is evaluated for various attributes, and decisions are made about the category or class it belongs in. Object categorization in odor processing is what allows us to recognize a familiar odor despite subtle changes in intensity or identity of its components, allowing

for stability in perception (Gottfried, 2010; Wilson & Sullivan, 2011).

Research done by Gregory Ashby and his associates over the last few decades has given important insight on category learning. According to Ashby, category learning can be explicit (decision process for categorization is easily verbalized) or implicit (decisions for categorization are internal and cannot be verbalized) (Smith et al., 2015). The attributes used to categorize stimuli can be either separable (visual shape, size, etc.) or integral (color, hue). Separable stimuli tend to be categorized explicitly while integral stimuli require implicit processes (Soto & Ashby, 2015). Thus, odor intensity would be considered separable and could be learned through explicit processes while odor quality would be considered integral, requiring a more implicit cognitive process. Ell, Ashby, & Hutchinson, (2012) evaluated whether category learning of colors could occur in an unsupervised (passive) setting. The results of that study indicated that separable attributes such as brightness could be learned without feedback; but hue, an integral attribute, required supervision in the form of feedback for perceptual learning.

1.3 Odor Mixture Analysis

Successfully developed methods can isolate, identify, and quantify volatile odor compounds in food products with precision and accuracy. Aroma extract dilution analysis (AEDA) (Grosch, 1993), charm analysis, and gas chromatography – olfactometry (GC-O) are commonly used methods for isolating and identifying odorous compounds (Dunkel et al., 2014). The compounds are typically quantified using stable isotope dilution assay (SIDA) or gas chromatography-mass spectrometry (GC-MS). Odor

Activity values, which measure the concentration of the compounds expressed in units of odor detection threshold, and flavor dilution (FD) factors (obtained from AEDA) are commonly used to identify the compounds that are considered the most odor-active (Dunkel et al., 2014). Although these methods help to narrow the list of hundreds of volatiles in a food product to only those that are perceivable, they cannot measure perceptual interactions between volatiles (including inhibition, suppression, and blending) (Grosch, 2001), and , therefore, sensory methods are necessary. One commonly used sensory method is omission testing.

1.3.1 A Review of Sensory Methods in Omission Studies

Omission testing is a frequently used method for determining which volatile compounds in a food product contribute to the perceived aroma. In omission testing a synthetic model of the target odor is created. The compounds chosen through instrumental/sensory evaluation are combined in the concentrations occurring naturally in the food product of interest in a matrix that simulates the food product. This model then is compared to the original product using sensory methods. Once a model is determined to be a good approximation of the original, individual compounds or groups of compounds are omitted to create new, partial models. To determine whether people can discriminate between the two, these partial models are compared to the full recombination model using sensory methods

Schieberle et al. (1993) completed a study evaluating the key odor compounds in five kinds of butter (sour cream butter, sweet cream butter, Irish sour cream butter, cultured butter, and farmer sour cream butter). They narrowed the list of volatile

compounds to the five with the highest FD factors in all five types of butter. Five trained panelists evaluated each butter sample by listing descriptive attributes (generated individually) and rating the overall odor intensity (1 weak, 2 medium, 3 strong). The Irish sour cream was chosen for omission testing. OAVs were calculated for the five compounds. Using the three compounds with OAVs greater than one, the researchers prepared a recombination model as well as three omission models (one compound omitted from each model) in a sunflower oil matrix. The five trained panelists compared the four models by describing the odor attributes of each model and scoring them on a scale of 1-3 for their similarity to the butter sample. The model containing all three compounds had a "strong similarity" to the butter sample. The omission models were all found to be less similar.

Guth and Grosch (1994) used omission testing to evaluate 15 compounds with OAVs greater than one in stewed beef juice. They combined the 15 compounds in a model matrix containing several non-volatile compounds, coconut oil, and water. Five expert panelists, trained with solutions of the 15 compounds, sniffed the original juice and five samples of the full mixture. They then evaluated the samples for overall intensity and quality in comparison to the original using a scaling method (0=absent, 1=weak, 2=medium, 3=strong). The mean similarity rating for the model mixture was 3.0 indicating a strong similarity. The panelists evaluated omission models in which one compound had been removed from each model in a similar manner. The mean similarity ratings for three omission models were 3.0 (a strong similarity to the original). Therefore, the researchers concluded that these three compounds were not important for

the aroma of the stewed beef juice. To verify these findings, the researchers prepared a final model containing the 12 (out of 15) compounds thought to contribute to the aroma. The five panelists compared the original meat juice with this final model and found a "strong" similarity. However, comparisons of similarity do not indicate that the aromas of the original stewed meat juice and the 12-compound model were indistinguishable.

Schieberle and Hofmann (1997) completed a study on the importance of 12 compounds having the highest flavor dilution (FD) factors in fresh strawberry juice. A model juice matrix was prepared to simulate the non-volatiles and color of the original juice. A team of six evaluated both the original juice and the 12-compound model by rating the intensities of eight odor qualities characteristic of the compounds of interest (green, leaf-like, caramel-like, fruity, sweet, sour, pungent, sweaty, and buttery) on a seven point scale (from 0.0 to 3.0). Schieberle and Hofmann, after plotting the mean intensity ratings on a spider web diagram, determined that the model "showed a clear similarity" to the original strawberry juice. Twelve omission models, prepared in the same matrix as the model omitting one compound from each, were evaluated by the same six judges by triangle tests with the full model as the control. No statistical analyses were reported (only raw data), but the researchers stated that omission of two of the compounds (five or six chose the correct sample) "led to a significant change" while four compounds (one or two chose correctly) had a "lower flavor impact." For triangle testing with six panelists and no repetition, statistical significance requires a minimum of five correct judgements (at a 0.05 probability level) or all six (at a 0.01 probability level) to determine that they could discriminate between the two samples (Lawless & Heymann,

2010). This would indicate that only two compounds (of the 12 compounds tested) contributed to the strawberry juice aroma. Odor profiles were completed on the two omission models that differed significantly from the full model as a strategy for comparing them to the original juice. Schieberle and Hofmann concluded that neither of the two omission models had a typical strawberry aroma.

Wagner and Grosch (1998) conducted a study using 19 compounds with OAVs greater than one in French fries prepared in palm oil. They believed that these compounds would make the greatest contribution to the French fry aroma. To test this assumption, Wagner and Grosch prepared a model of the 19 compounds in an oil matrix. Five trained and experienced panelists created a flavor profile of the model by rating the intensity of six attributes (earthy, deep fried/fatty, boiled potato, caramel, malty, coconut-like) on a six point category scale (0 = absent, 3.0 = strong). The researchers compared the mean values and concluded that the flavor profile was "close to that of the corresponding French fries." The same five panelists participated in omission testing. Triangle tests were used to determine if the panelists could discriminate between the full model and 19 partial models; nine omitting one compound, 10 omitting two or more compounds. In addition, Wagner and Grosch asked the panelists to rate the intensity of the difference between the full model and the omission model if a difference was found. Significant results were not obtained for eight of the nine models in which single compounds had been omitted (four or less panelists chose correctly, intensity ratings were 1.5 or less) indicating that the omission of these compounds did not make any perceptible difference. Interestingly, none of the panelists could discriminate between

the full model and the model with the compound methional omitted despite methional having the second highest OAV (OAV = 3915).

Reiners and Grosch (1998) studied odor compounds in virgin olive oil from Spain, Morocco, and Italy. After determining the "potent odorants" in each olive oil using quantitative results and OAVs, they prepared three recombination models in a plant oil matrix. Nine experienced panelists were trained using reference samples prior to rating the intensities of seven attributes (pungent, fruity, green/leaf-like, fatty, apple-like, black currant-like, and black olive-like) in the three oil samples and their models using a category scale (0 to 3.0). The similarity of each model to its original oil sample was also rated. Omission testing was done by removing individual or groups of compounds from all three recombination models for a total of 15 partial models lacking one compound and eight lacking groups of compounds. The similarities of the partial models to the corresponding full models were rated using a seven-point scale (0 = "no similarity", 3.0 = "identical with the complete model") by the same nine panelists. Seven of the 15 partial models resulted in mean similarity ratings greater than 2.1 (a strong similarity to the full model), suggesting that the omitted compounds made little contribution to the perceptible odor. However, the results of similarity ratings do not indicate whether panelists could distinguish between the full olive oil models and the models omitting one compound.

Czerny el al. (1999) studied 27 odor compounds in roasted Arabica coffee. Ten experienced panelists evaluated four bases for aroma models and a coffee sample by rating the intensity of four attributes (sweetish/caramel, earthy, roasty/sulfurous, and smoky) using a scaling method (0 = no similarity, 3 = identical with the coffee). The

panelists also rated the similarity of the four models to the coffee sample using the same scale. The research team determined that an oil and water base was the most similar to the original coffee based on the mean value of the results. The oil and water base was then used for the preparation of 20 partial models in which individual or groups of compounds were omitted (nine omitted one compound, 11 omitted two or more). The researchers used "duo tests" (testing 16 models) and triangle tests (testing 18 models) to determine if panelists could discriminate between the full model and the partial models. The triangle tests were completed in duplicate over two sessions. The results indicate that removing three of the nine compounds tested individually resulted in detectable differences.

A second study was done by the same group (Mayer, Czerny, & Grosch, 2000) investigating which of 25 volatile compounds was responsible for the perceived aroma of a coffee beverage brewed from roasted Arabica coffee beans. Ten experienced panelists created an attribute profile of a model of the 25 compounds in the concentrations found in the original coffee samples by rating the intensities of four attributes (sweetish/caramel-like, earthy, roasty/sulfurous, and smoky) on a seven-point scale (0 = no similarity to 3 = very strong). The research group determined that the model was a good approximation of the original aroma despite finding that the intensity of some attributes was significantly different between the original coffee and the model. The researchers prepared 24 omission models: 13 models omitting one compound, 11 omitting groups of compounds. They used triangle tests to determine if panelists could discriminate between the full model and each of the omission models. Each omission model was evaluated by 10

panelists twice, using triangle tests. Mayer el al. found that there was a significant difference (11 or more correct answers out of 20 possible) between the full model and 11 (three omitting one compound) of the 24 partial models. These results indicate that removing 10 of the 13 compounds tested individually produced no detectable difference.

Ferreira et al. (2002) used omission testing to evaluate 22 (all of the compounds with OAVs greater than 0.5) of the 38 aroma compounds identified in Grenache rose wines. Eleven trained panelists compared the full reconstitution model, containing all 22 compounds of interest in a synthetic wine matrix, to the original wine using triangle tests. This method was followed by asking panelists to complete an extent of difference scale (0 = null, 1 = slightly different, 2 = quite different, and 3 = full different) and describe the level of difference between the original wine and the model. The results of the triangle test were not reported but were said to be significant indicating that panelists were able to discriminate between the wine and the model. The collective distance rating was 0.4 and the description of the difference between the two was reported as a "slight difference in intensity." The researchers then used the reconstitution model for omission testing. The panelists evaluated the full reconstitution model and 21 omission models (20 omitting one compound and one omitting two) using triangle tests. If significant differences were found, the panelists completed an extent of difference scale and described the differences between the samples. After triangle testing, four of the omission models (three omitting one compound) were significantly different from the full mixture with scores on the extent of difference scale high enough to allow for describable differences between the omission model and the full model. Nine more mixtures were significantly different from

the full mixture following triangle testing but the difference was not enough to be describable. The results of triangle testing on eight omission models were not significant. Ferreira and his associates concluded that the eight compounds "could be removed from the model without any noticeable change in its aroma." However, they never tested this claim by comparing the model with these eight compounds removed from the model with all 22 compounds.

Escudero et al. (2004) studied 53 odor compounds in wine from Maccabeo. Three reconstitution models were prepared in a de-aromatized wine matrix: one containing all 53 compounds, one containing all compounds with OAVs greater than 0.1, and one containing all compounds with OAVs greater than 1.0. Between 16 and 18 panelists compared the models, the original wine, and the dearomatized wine matrix using triangle tests followed by duo-trio tests. Significant differences were not found among the three models - but, despite these results, the reconstitution model containing all 53 compounds was used for omission testing. All of the compounds with OAVs greater than one (21) were omitted individually from the full model for omission testing. Between 14 and 16 panelists evaluated the reconstitution model and omission models using triangle tests. If triangle testing results were significant, the panelists completed duo-trio tests with the original wine as a reference. The triangle tests resulted in significant differences between the reconstitution model and seven of the 21 omission models. Only one of the seven omission models was significantly different following duo-trio testing. The researchers were disappointed with these results, but one has to

question the validity of a method in which only one of 53 compounds is found to be an important contributor to the perceived aroma of the mixture.

Steinhaus and associates (2009) used omission testing to evaluate 13 volatile compounds, all with OAVs greater than one in pink guava fruit, to determine which actually contributed to the perceived aroma. The researchers combined the 13 compounds in water for aroma reconstitution. A sensory panel of 15 to 20 trained panelists completed an aroma profile test of both the original fruit and the reconstitution model by rating the intensities of eight descriptors (fruity, grapefruit, fresh, grassy, seasoning-like, metallic, flowery, and caramel/sweet) on a seven-point scale (0 = not detectable, 1 – weak, 2 = moderate, and 3 = strong). The results were plotted on a spider web graph, and the researchers concluded that the reconstituted model was close to that of the original guava fruit by comparing the plots. The researchers prepared omission models by removing each of the 13 compounds once from the reconstituted model. Between 14 and 19 panelists evaluated the omission models using triangle tests. The results for eight of the 13 compounds were significant indicating that panelists could discriminate between the 13-component mixture and the mixture without that compound. The results for five were not significant. The researchers determined that, due to the complexity of the mixture, these compounds had been suppressed. They concluded that the omission testing had "clearly elucidated the key aroma compounds of guava fruits." However, a mixture of the compounds found significant never was compared to the original reconstituted model or the guava fruit to test that conclusion.

Burdack-Freitag and Schieberle (2012) conducted research on 19 odor compounds in raw hazelnuts and 25 compounds in roasted hazelnut paste, all with the highest FD factors (as determined in a previous study). Following separation and quantification they calculated OAVs and prepared two recombination models by combining all of the compounds with OAVs greater than one in raw hazelnuts (15) and roasted hazelnut paste (23) in a sunflower oil matrix. The researchers utilized a panel of 12 trained and experienced staff members to rate the intensities of 11 odor attributes for raw hazelnuts (fruity / nutty, fatty, green / citrus-like, earthy, sour, popcorn-like, flowery, bell-pepper-like, malty, potato-like, and phenolic) and 13 attributes (fruity / nutty, earthy, fatty, malty, sweet, buttery, spicy, potato-like, sour, green / citrus-like, popcorn-like, coffee-like, and phenolic / smoky) for roasted hazelnuts using a seven-point linear scale (0.0, 0.5, 1.0, …, 3.0) in triplicate. The mean values were graphed in a spider web diagram and the researchers compared the aroma profiles of both models to the original samples. They determined that the models were "a good similarity" to the raw and roasted products. Burdack-Freitag and Schieberle prepared 13 omission models: seven in which individual compounds were omitted and six in which "groups of compounds exhibiting the same odor quality" were omitted from the full roasted hazelnut paste model. Fourteen panelists compared the omission models to the recombination model using triangle tests. Unlike the other omission studies reviewed, the researchers reported a significant difference for all 13 models (10 – 14 chose the correct model).

Tokitomo and associates (2014) used omission testing to evaluate six compounds with the highest OAVs in fresh pineapple. A model mixture of the 12 compounds

identified by AEDA and SHA to be the most odor-active in fresh pineapple was prepared in distilled water. Working with 15 panelists (trained by evaluating solutions of four of the 12 compounds), the researchers used a seven-point scale (0 = absent, 6 = strong) to evaluate each of seven odor qualities (sweet, citrus-like, fresh, fruity, green and grassy, woody, and pineapple-like) in both the original fruit and the full mixture. The researchers compared the results on spider web plots and determined that there were no "significant" differences between the pineapple juice and the model mixture for any of the seven odor qualities. They prepared partial models in which one of the six compounds having the highest OAVs had been omitted. The same 15 panelists compared the partial mixtures with the 12-component model using triangle tests. The results for two of the six omission models were significantly different, indicating that panelists could discriminate between the omission models and the full model. The results for four were not significant.

Gao and associates (2014) evaluated 27 compounds having an OAV greater than or equal to one in Chinese liquor. They created an aroma recombination model by combining all 27 compounds in a hydroalcoholic solution. Ten trained (previously) panelists evaluated eight attributes (fruity, floral, acid, earthy, alcoholic, grassy, mushroom, and coconut) in both the original liquor and the recombination model by rating the intensity of the attribute on a six point scale (0 = none, 1 = very weak, 2 = weak, 3 = moderate, 4 = strong, and 5 = very strong). Gao and his team plotted the results on a spider web diagram and found the model to be a good match to the original product in all eight attributes. They prepared 15 omission models from the recombination model: eight removing individual compounds (with the highest OAVs),

six removing groups of compounds. The same 10 panelists then participated in omission testing using triangle tests. Seven of the 15 models were found to be significantly different from the full model indicating that panelists could detect a difference between the samples. Four of these were models with one omitted compound and three with omitted groups of two or more compounds. Eight of the 15 models were not significantly different. Four of these were models with one omitted compound.

Pavez and associates (2015) used omission testing to evaluate two volatile compounds recently identified in red wine. They prepared five models for sensory analysis: four models with 28 major odorants as determined in previous research as well as varying concentrations of the two new compounds and one model omitting the new compounds. All models were prepared in a wine matrix. Twenty-four experienced panelists performed a series of triangle tests to determine whether they could discriminate between the samples with the new compounds (at any of the four concentrations), and the samples without the new compounds. The panelists were unable to discriminate between the models with the new compounds (at any concentration) and the model omitting the new compounds. The Pavez team determined that the newly identified compounds were not important in the overall flavor profile of the wine.

A study conducted by Kiatbenjakul and others (2015) investigated the odor compounds in male giant water bugs. They prepared an aroma model by combining 27 volatile compounds, all the compounds in salted boiled male giant water bugs with flavor dilution factors above nine. The researchers recruited 11 panelists who reported familiarity with the consumption of the water bugs. The panelists created an aroma

profile of the product using scaling (0 = none, 15 = extremely strong) to rate the intensities of nine attributes (cooling, pineapple-like, fruity, green apple-like, banana-like, floral, salty, solvent-like, and fishy). The results were plotted on a spider web design. Eight omission models (seven omitting one compound and one omitting a group of compounds) also were prepared. Thirty panelists evaluated the omission models using the R-index method. Each of the omission model samples was compared to a control (the full model) and ranked from most similar to least similar. Only one omission model (omitting one single compound) was found to be significantly different from the full model. The researchers determined that the compound omitted from this model was "the only character–impact odorant in the model," and its omission model was chosen for further study. Six models omitting one compound were not significantly different from the full model indicating that panelists could not perceive a difference when the compounds were omitted.

Most recently, Delime et al. (2016) used a synthetic strawberry aroma to compare both orthonasal and retronasal OAVs with the results of sensory omission testing. The synthetic blend consisted of nine odor compounds in propylene glycol. The researchers prepared nine partial models omitting one compound from each model. They recruited 100 panelists each for orthonasal and retronasal testing. The panelists used same-different discrimination tests to compare the full model with each of the nine omission models. Following each test the panelists were asked to rate the sureness of each answer on a four-point scale (very unsure, unsure, sure, and very sure) Delime and associates analyzed the data using signal detection theory to measure d'. Following orthonasal

omission testing, significant results were found for two of the nine models (both with d' values greater than one) indicating that the omission of seven of the compounds was not detectable by the panelists.

In studies evaluating the odor compounds in Sauvignon Blanc wine (Benkwitz et al., 2012), muskmelon juice (Pang, Chen, Hu, Zhang, & Wu, 2012), and Shiraz wine (Mayr et al., 2014) omission testing was completed using descriptive analysis techniques. The participants evaluated odor attributes in original samples, recombination models, and omission models (omitting both groups of compounds and individual compounds) using line scales. Principle component analyses were performed in all three studies to compare the aroma profiles of the full model and the omission models. The results were used to determine which compounds had the greatest effect on specific attributes in the original product. Detectable differences between the partial models and full models were not apparent from the principal component analysis.

All of the omission testing research cited here follows the assumption that omission testing works, and that it is a valid method for assessing the contributions of odor compounds to the overall perceived aroma. However, to the best of our knowledge, omission testing has never been validated. In most of the omission testing studies the full recombination models did not include all of the volatiles detected in the product, but were made using only the compounds deemed important based on OAVs (Gao et al., 2014; Reiners & Grosch, 1998; Wagner & Grosch, 1998), FD factors (Kiatbenjakul et al., 2015), or both (Burdack-Freitag & Schieberle, 2010; Ferreira et al., 2002; Guth & Grosch, 1994; Schieberle et al., 1993; Steinhaus et al., 2009; Tokitomo et al., 2014),

neglecting contributions of compounds that were in concentrations below their odor threshold but may still play a role in the perception of the mixture. What's more, the study conducted by Schieberle and associates (1993) on the odorants in stewed beef juice is the only example we could find in which a model of the odorants found to be responsible for the aroma as a result of omission testing were recombined and compared to the original product.

The purpose of discrimination testing (such as triangle testing) in sensory evaluation is to determine whether two samples differ perceptually (Lawless & Heymann, 2010). When the statistical results following omission testing indicate that panelists are not able to discriminate between the full model and the omission models there are several possible explanations only one of which is that the compound simply does not contribute to the perceived aroma.

The contributions of a volatile compound to the perceived aroma may go undetected in omission testing if the panelists have not had adequate training for the task. Many of the omission testing studies we reviewed reported using trained or experienced panelists for sensory evaluation of omission models. However, most of the studies were vague about what constitutes training or experience (Czerny et al., 1999; Mayer et al., 2000; Schieberle et al., 1993; Steinhaus et al., 2009). Some indicated that panelists were experienced because they were chosen from staff or laboratory members and were, therefore, familiar with the product (Burdack-Freitag & Schieberle, 2010; Ferreira et al., 2002; Gao et al., 2014). In the study on red wine Pavez and associates (2015) reported that the panelists were experienced with wine tasting. Some trained the panelists for

descriptive analysis using reference solutions but not specifically for discrimination testing (Gao et al., 2014; Reiners & Grosch, 1998; Wagner & Grosch, 1998). Only two studies reported training panelists with either individual compounds (Guth & Grosch, 1994) or solutions of the compounds used in omission testing (Tokitomo et al., 2014). Research has shown that people are better able to discriminate odor quality following training or experience (Jehl et al., 1995) (Rabin, 1988). What's more, research performed by Sinding and associates (2013b) suggest that complex odor mixtures that blend into odor objects have unique characteristics separate from their individual components. Experience or training with only the individual components or reference solutions may not adequately prepare panelists for discrimination tasks involving a complex odor mixture. Experience with the product may also not be adequate because omission testing does not involve discriminating from the original food product - but, rather, a model. The model may be similar to the original, but it is not exactly the same and may not trigger the same pattern of activity in cortical processing. Therefore, the panelists may not have been able to discriminate between full and partial mixtures even if the missing component was an important odor compound.

A volatile compound that has an important role in the perceived aroma of a food product may not be detected in omission testing because there were too few panelists for a significantly powered experimental design. Statistical power is a measure of the probability of finding a true difference between two samples if a difference exists. One factor that affects the level of power in a test is sample size or, in the case of discrimination testing, number of panelists. Too few panelists results in a low powered

study and a high probability of not finding significant differences when differences exist. Many of the omission studies we reviewed reported using triangle tests with between five and 24 panelists (Burdack-Freitag & Schieberle, 2010; Czerny et al., 1999; Ferreira et al., 2002; Gao et al., 2014; Mayer et al., 2000; Pavez et al., 2015; Steinhaus et al., 2009; Tokitomo et al., 2014; Wagner & Grosch, 1998). The power of a test can be calculated using Equation (1.1) where x is the number of correct responses, n is the total number of panelists, p_d is the proportion of distinguishers, p_o is the probability of a correct guess (0.33 for triangle tests (Lawless & Heymann, 2010)), and B is the binomial distribution probability (Meilgaard, Civille, & Carr, 2016). A triangle test with five panelists would require at least four correct responses for significant results ($\alpha = 0.05$), and the power of this test would be 0.23. This means that there is only a 23% chance of finding a difference between the full mixture and partial mixtures if a difference exists. A triangle test with 24 panelists would require 13 or more correct responses at the same significance level, and the power of this test would be 0.54, or a 54% chance of finding a difference if one exists. The desired proportion of distinguishers, or effect size, for these studies was not reported so we calculated power using an arbitrary value of 0.30.

$$Power = 1 - B(x; n, (p_d + p_o * (1 - p_d))) \quad (1.1)$$

One final reason why panelists may not be able to discriminate between the full mixture and mixtures with omitted compounds, even important odor compounds, is that the mixtures may have blended or fused into a unique odor object. Odor objects are the result of synthetic processing of complex odor mixtures in the piriform cortex (Jinks & Laing, 2001; Stevenson, 2001). Patterns of receptor activity are processed and stored for

later retrieval. But incoming stimuli is never a perfect match. When incoming patterns of activity are similar enough to one stored in memory, the brain uses pattern completion to promote perceptual stability. Pattern completion suggests that the brain may replace the missing compound in complex mixtures (Wilson & Sullivan, 2011).

Omission testing as a method for determining the important compounds in the overall perceived aroma of foods sounds logical. If an odor compound contributes to the perceived aroma people would notice if it is removed. Clearly, if panelists can discriminate between the full model and one with an omission, the omitted compound is important to the perceived aroma. However, the opposite results may not be as meaningful.

Hongsoongnern (2003) used omission testing to determine whether panelists could learn to discriminate between a five compound mixture (with compounds at a similar perceived intensity) and the same mixture with any one of the five compounds removed. She hypothesized that learning would be implicit; panelists would learn to discriminate between the full mixture and ones with omitted compounds but they would not be able to verbalize the criteria used for discriminating.

The study began with intensity matching. Seventeen panelists used an unstructured line scale (labeled "Much less than standard" on the far left, "Exactly same as standard" in the middle, and "Much stronger than standard" on the far right) to rate the intensities of four compounds at five different concentrations each in comparison to the perceived intensity of a fifth compound used as the standard. Hongsoongnern analyzed the results for each compound for each individual panelist using regression analysis with

concentration as the independent variable and intensity rating as the dependent variable. The concentrations matching the intensity of the standards were calculated from the linear or quadratic equations for the trend lines. The calculated concentrations were then used to prepare individualized sample mixtures for the panelists for omission testing

For omission testing, the panelists evaluated 20 samples (10 full mixture samples, 10 partial mixture samples with one compound removed from each sample and each compound removed from two samples) in each of 20 sessions using the A-Not-A method with verbal feedback. At the end of 20 sessions, each panelist rated the difficulty of the task on a 10 point scale (1 = "Extremely easy", 10 = "Extremely difficult") and described the criteria used to make decisions during testing. Hongsoongnern analyzed the data for each panelist and the combined data for all panelists using linear regression with session as the independent variable and the number of correct responses as the dependent variable. The slope of the regression line was used to evaluate learning. A positive slope to the line would indicate an improvement in performance and thus that learning had occurred.

Initially, the average number of correct responses for the group, 10.64 out of a possible 20 (53.2%) (1.2), was barely above chance (50%) indicating that the panelists were not able to easily discriminate between the full mixture and mixtures with one missing component. A positive slope to the regression line indicated that performance improved but only at a rate of 0.11 correct responses per session (0.005%). Twenty sessions and 400 trials later, the group's performance, 12.84 correct responses (64.2%), was still not much higher than chance level. Individual results indicated that three of the

nine panelists who completed the study had slopes around zero indicating that they never did learn to discriminate between the full mixture and the partial mixtures. The slopes for the six remaining panelists ranged from 0.0910 to 0.2316 for an increase in correct responses after 20 sessions of 1.82 (0.09%) to 4.63 (23%).

$$Total\ Correct = 10.64 + 0.11 * Session \quad (1.2)$$

The average difficulty rating for the 11 panelists was 7.0. The panelists were not able to verbalize a rule for decision making, so Hongsoongnern concluded that learning was implicit.

1.4 Signal Detection Theory

Signal detection theory (SDT) is a method frequently used to measure discriminability. In SDT a theoretical decision-making model is used to determine whether panelists can discriminate between samples and how difficult the task is. Two types of samples are presented: samples of interest (signals), and distractors (noise). Panelists are asked whether the sample is or is not a signal, and their responses are recorded in a response matrix (Figure 1.1). This matrix allows for four possible outcomes. If the panelist is given a sample that is a signal and the panelist's response is "yes," then the response is counted as a "hit". If the response is "no," then it is counted as a "miss." If a noise sample is presented and the response is "yes," then the response is counted as a false alarm, and if the response is "no," then it is a correct rejection (Anderson, 2015). Performance is measured by calculating a hit rate (probability of

responding “yes” when a signal is presented) and a false-alarm rate (probability of responding “yes” on a distractor). (Abdi, 2007; Anderson, 2015; Heeger, 1998)

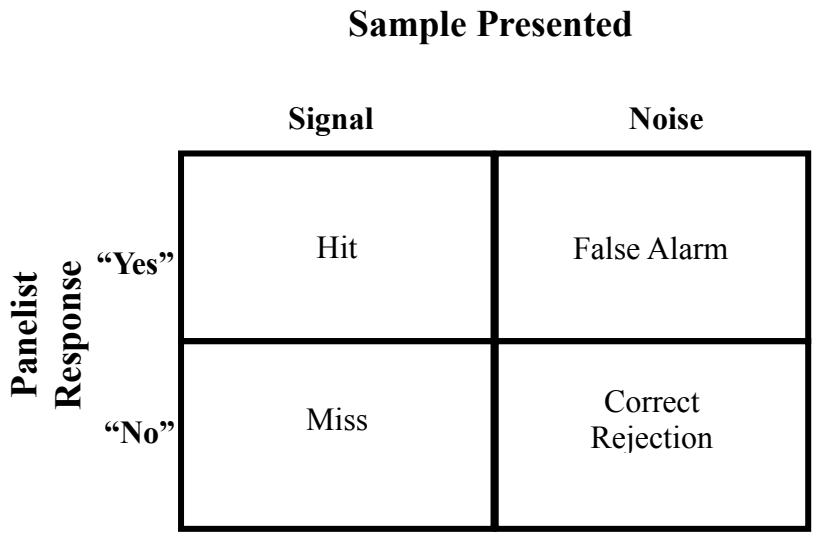


Figure 1.1: A signal detection theory response matrix.

In SDT a probability of occurrence curve is used to illustrate the ability of panelists to discriminate between samples (Figure 1.2). The horizontal axis is a measure of “internal response.” This may be a measure of sensory evidence, neural activity, or some underlying decision variable. The vertical axis is a measure of probability (Heeger, 1998). The panelists’ responses to both the signal and noise samples are assumed to follow a normal (Gaussian) distribution (Anderson, 2015). The noise distribution accounts for responses to the distractor samples as well as any other sources of noise including both internal (neural activity) and external (environmental stimuli, variations in samples, etc…). Because much of the internal and external noise is still present, the

distribution curve for responses when signals are presented is shifted to the right on the x-axis (Heeger, 1998).

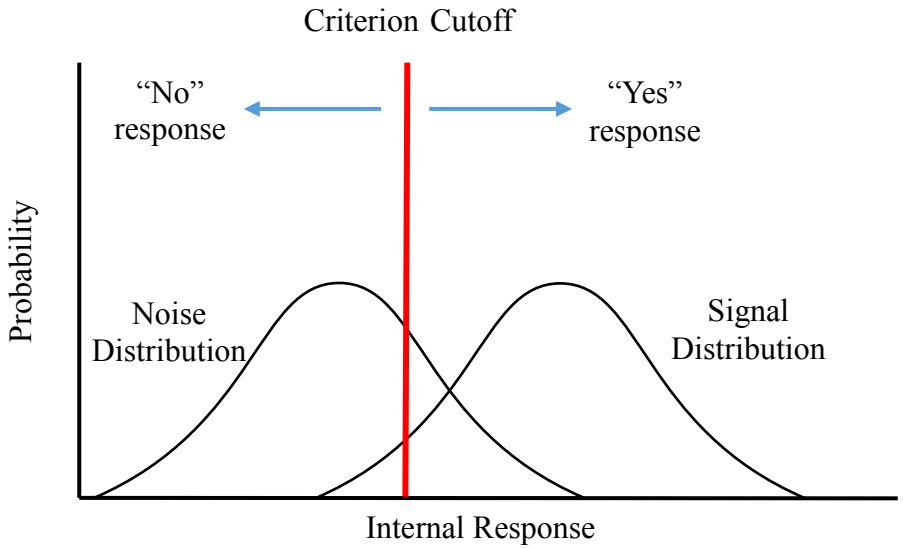


Figure 1.2: A signal detection theory probability of occurrence curve with the signal and noise distributions. The horizontal axis is a measure of internal response to a stimuli and the vertical axis is a measure of probability. A criterion cutoff is established by the panelist during the decision making process.

When presented with a sample the panelist is forced to make a decision. The panelist will develop some sort of decision making variable or criterion cutoff (Anderson, 2015). If there is enough evidence to support the cutoff, the panelist will respond "yes" when asked if the sample is a signal. If there is not enough evidence to meet this cutoff, the response is "no." If the panelist tends to be very conservative in decision making and frequently responds "no," this cutoff line will move to the right in the probability of occurrence curve. If the panelist is liberal and responds "yes" to everything, the line moves to the left. However, SDT estimates a difference independently of where the

cutoff is placed. One of the benefits of SDT is that it removes response bias (e.g. conservative v liberal) from the measure of discriminability (Abdi, 2007).

The panelist's ability to discriminate between the signal and noise samples is measured by calculating the discriminability index (d'): a measure of the distance between the center points of the two distributions in the probability of occurrence curve (Figure 1.3). The easier it is to discriminate between the samples, the farther the two curves are apart, and the greater the value of d' (Figure 1.4) (Abdi, 2007). A d' of 0 would indicate that the two curves completely overlap, and the samples are indistinguishable. As d' increases, panelists are better able to distinguish between the samples, and the curves separate until a d' of four where there is no similarity between samples (Abdi, 2007).

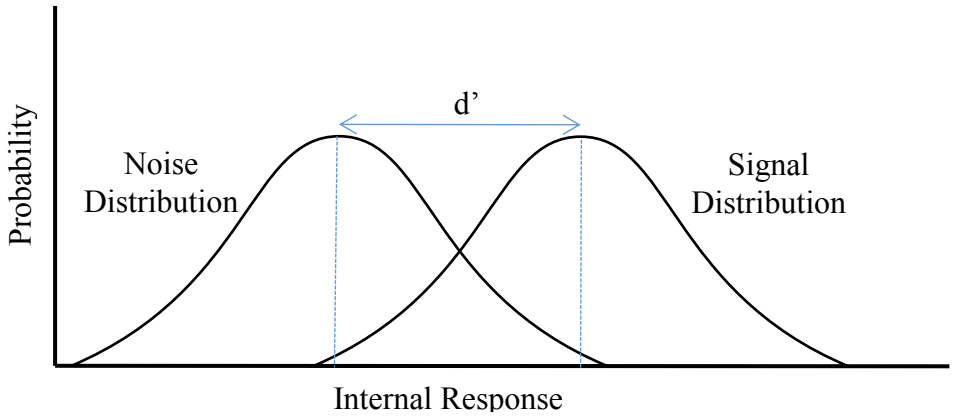


Figure 1.3: A diagram of the discriminability index (d'). The discriminability index is a measure of the distance between the center points of the noise and signal distributions in the probability of occurrence curve.

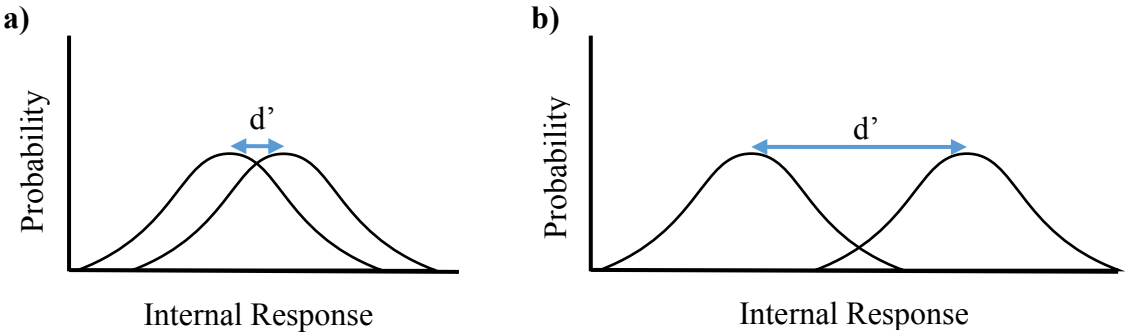


Figure 1.4: a) A signal detection theory probability of occurrence curve with high overlap and a small value for d'. **b)** A signal detection theory probability of occurrence curve with low overlap and a large value for d'.

The location of the criterion cutoff is used to determine the distance between the two curves. The location of the criterion cutoff can be determined by the hit rates and false-alarm rates obtained from the response matrix. The proportions of hits and false alarms correspond to areas under the normal curves. The probability of responding "yes" when a signal is presented (hit rate) corresponds to the proportion of the signal distribution curve to the right of the criterion cutoff. The probability of responding "yes" when a distractor sample is presented (false-alarm rate) corresponds to the proportion of the noise distribution curve to the right of the criterion cutoff (Figure 1.5) (Lawless & Heymann, 2010). The z-scores corresponding to these areas can be obtained or calculated. An estimation of d' is calculated using (Equation (1.3) (Lawless & Heymann, 2010).

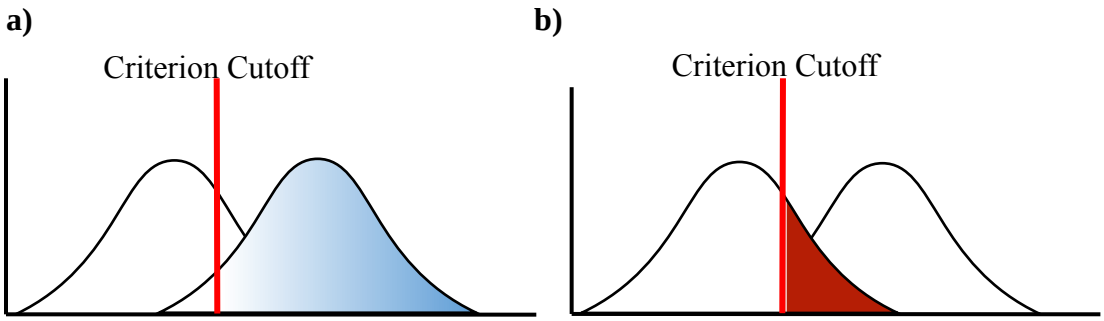


Figure 1.5: a) A signal detection theory probability of occurrence curve. The hit rate corresponds to the proportion of the signal distribution curve to the right of the criterion cutoff (shaded portion). **b)** A signal detection theory probability of occurrence curve. The false alarm rate corresponds to the proportion of the noise distribution curve to the right of the criterion cutoff (shaded portion) (Lawless & Heymann, 2010).

$$d' = z_{hits} - z_{false\ alarms} \tag{1.3}$$

2 Objectives and Hypotheses

2.1 Phase 1: Intensity Matching

Objective 1.1: Screen panelists for their ability to accurately discriminate differences in intensities among different concentrations of each odor compound.

Objective 1.2: Obtain the concentrations of five compounds that match in perceived intensity for each panelist.

2.2 Phase 2: Omission Testing

Objective 2.1: Determine whether panelists could learn to discriminate between a five-compound odor mixture with all compounds at the same perceived intensity and the same mixture with one of the five compounds removed.

Hypothesis: Panelists will not be able to discriminate between the full five-compound mixture and mixtures when any one of the five compounds is removed, and they will not learn to discriminate between them over time.

Objective 2.2: Determine whether panelists could learn to discriminate between a five-compound odor mixture with all compounds at the same perceived intensity and the same mixture with two of the five compounds removed.

Hypothesis: Panelists will not be able to discriminate between the mixtures initially but will learn to discriminate between them when two compounds are removed.

Objective 2.3: Determine whether there is a difference among panelists in their ability to discriminate between a five-compound odor mixture with all compounds at the same perceived intensity and the same mixture with one or two of the five compounds removed.

Hypothesis: There will be no difference among panelists' ability to discriminate between odor mixtures when any one or two of the compounds are removed.

Objective 2.4: Determine whether panelists' ability to discriminate between a five-compound odor mixture with all compounds at the same perceived intensity and the same mixture with one or two of the five compounds removed was dependent on the identity of the omitted compound.

Hypothesis: The identity of the omitted odor compound will not make a difference in panelists' ability to discriminate between the mixtures when one compound is removed from a blended mixture.

3 Phase 1: Intensity Matching

3.1 Participants

We recruited 35 panelists via email (Appendix A) from a list of potential subjects maintained by the Sensory Center at the University of Minnesota. Interested panelists were asked to complete a survey constructed in and maintained by Qualtrics LLC© (Provo, Utah) (Appendix B). Panelists were chosen based on availability.

We obtained prior approval for this study from the University of Minnesota's Institutional Review Board. Panelists read a consent form (Appendix C) and gave verbal consent before beginning the study. As compensation, panelists received $5 per completed session.

3.2 Intensity Matching Samples

Our interest for this study was to evaluate the effectiveness of n-1 omission testing in mixtures in which the odors have blended or fused. To ensure that the mixtures formed a blend of compounds and no individual compound dominated, we first needed to obtain the concentration of each compound that matched in perceived intensity for each panelist. The five compounds we used were acetylpropionyl, (2,3-pentanedione), butyric acid, furaneol, δ-decalactone, and methional. These compounds all occur naturally in Cheddar cheese or are similar to compounds found in Cheddar cheese. We purchased the compounds from Sigma-Aldrich™, Inc. (Milwaukee, WI).

We chose acetylpropionyl as the standard for comparison at a concentration of 5 ppm. This concentration was chosen after informal testing showed it to be similar to that of a six on the 12-point butanol scale (Sweeten, McFarland, Sorel, Gauntt, & Redded, 1984). We prepared the standard solution in a 250 mL volumetric flask by diluting the compound with distilled water from Premium Waters Inc. (Minneapolis, MN). We transferred approximately 20 mL of the solution to 50 mL glass jars with lids. We then labeled the jars "Standard".

We prepared samples of butyric acid, δ-decalactone, furaneol, and methional at five different concentrations (Table 3.1). We first prepared stock solutions of each compound in 1,000 mL volumetric flasks. We used the stock solutions to prepare 250 mL of each sample solution at each concentration. We prepared individual test samples by transferring approximately 20 mL of sample solution into 60 mL lidded glass jars. The jars were coded with random 3-digit numbers, generated by SIMS™ Sensory Software

(Berkeley Heights, NJ), for identification and to prevent response bias (Figure 3.1). Samples were prepared 24 to 48 hours in advance and stored in a refrigerator. We removed samples from the refrigerator approximately one hour before testing to allow them to come to room temperature.

Table 3.1: Concentrations (ppm) of butyric acid, δ-decalactone, furaneol, and methional prepared for intensity matching.

Compound	**Concentration (ppm)**				
Butyric Acid	0.1	50	100	150	200
δ-Decalactone	0.1	25	50	75	100
Furaneol	0.1	50	100	150	200
Methional	0.001	0.01	0.1	1.0	10

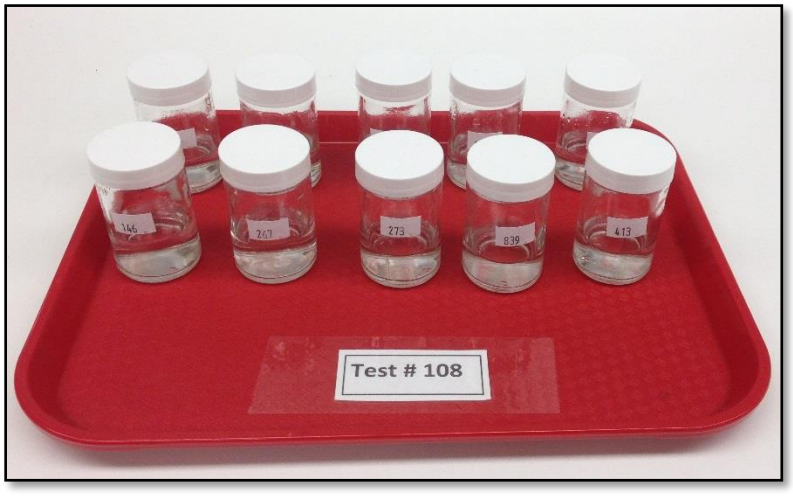


Figure 3.1: Furaneol intensity matching samples. All of the samples on one tray were samples in one test set with two samples each at five different concentrations.

We grouped the samples on trays in sets with a single compound per set and one set per tray. A set contained two samples of each of the five different concentrations for a total of 10 samples. To ensure that the proper trays and samples were presented to each panelist in the proper order, we color coded the trays and labeled the trays with 3-digit random numbers (also generated by SIMS Sensory Software), signifying the test set they represented (Figure 3.2). Because we were unable to obtain a concentration matching the intensity of the standard for all panelists using these sample sets, we prepared a second set for retesting using the wider range of concentrations (Table 3.2).

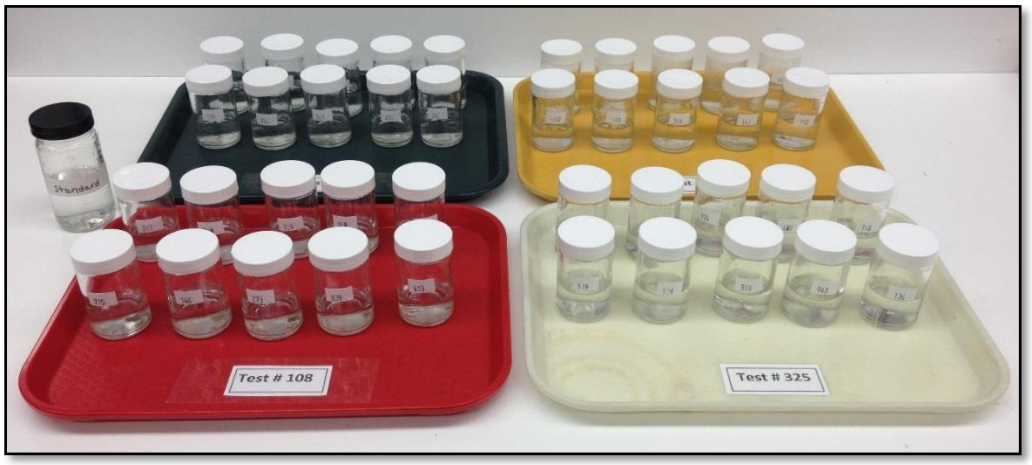


Figure 3.2: The four sets of samples used for intensity matching.

Table 3.2: Concentrations (ppm) of butyric acid, δ-decalactone, furaneol, and methional prepared for intensity matching retesting.

Compound	Concentration (ppm)				
Butyric Acid	0.01	1.0	100	125	250
δ-Decalactone	0.01	10	75	150	200
Furaneol	1.0	75	150	250	500
Methional	0.001	0.01	0.1	1.0	10

3.3 Intensity Matching Sessions

The first phase of the study consisted of one 30-minute session for each panelist. A session required four separate tests, one for each of the four sets of compounds described above. Each panelist needed to log into the software for each test

Each panelist arriving for the session received a panelist test order slip (Figure 3.3) containing information necessary for logging into the SIMS Sensory Software on booth computers. The information included a panelist identification number and a list of the four test codes in the order in which the tests should be taken. The test order varied among panelists and was determined using a Latin square design (Williams, 1949).

You are panelist #_____

Today you will be completing 4 tests. The tests must be completed in the order listed. Log in using your panelist number and slide this card through the window. You will need to log back in between tests.

Test #:

472
683
108
325

Please slide your tray & this card through the window when you have finished each test.

Figure 3.3: An intensity matching panelist test order slip.

For each test, we presented the panelist with an acetylpropionyl standard and 10 samples of one compound. The 10 samples consisted of two each of five different concentrations (Table 3.1 or Table 3.2) on a color-coded tray labeled with the test code. Each test was directed and scored using SIMS Sensory Software on computers in sensory booths.

We instructed the panelists to smell the standard and remember its intensity. We then instructed the panelists to smell each sample and compare the intensity of the sample to that of the standard. We reminded them to focus on the intensity of the sample and standard, not the individual odor characteristics. We instructed them to rate the intensity compared to that of the standard by placing a mark on a 150 mm line scale with endpoints labeled “Much less intense” and “Much more intense” and the center labeled

"Same intensity as standard" (Figure 3.4). The samples were presented in a balanced order, controlled by the software, to prevent order and carryover effects (Lawless & Heymann, 2010). The panelists were reminded to smell the standard before each sample. They were to wait 30 seconds between each sample and two minutes between each test to minimize adaptation. The wait time was controlled by the software. See Figure 3.5 for an example of the instructions.

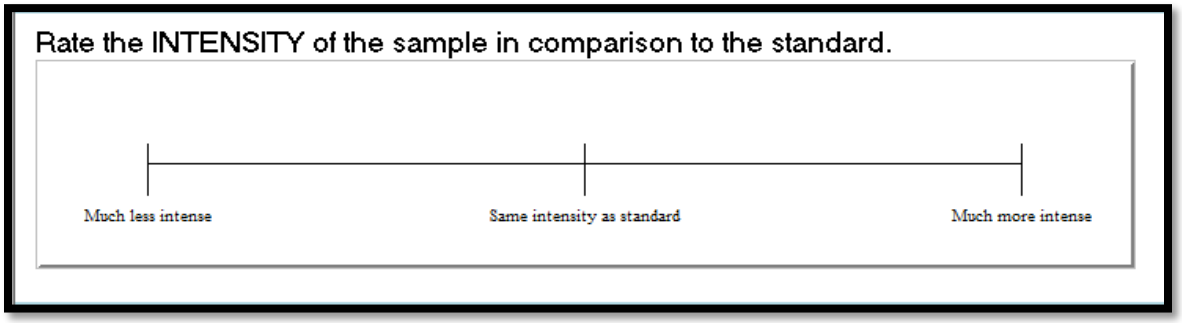


Figure 3.4: The line scale used by the panelists to rate the intensity of each sample in comparison to the standard.

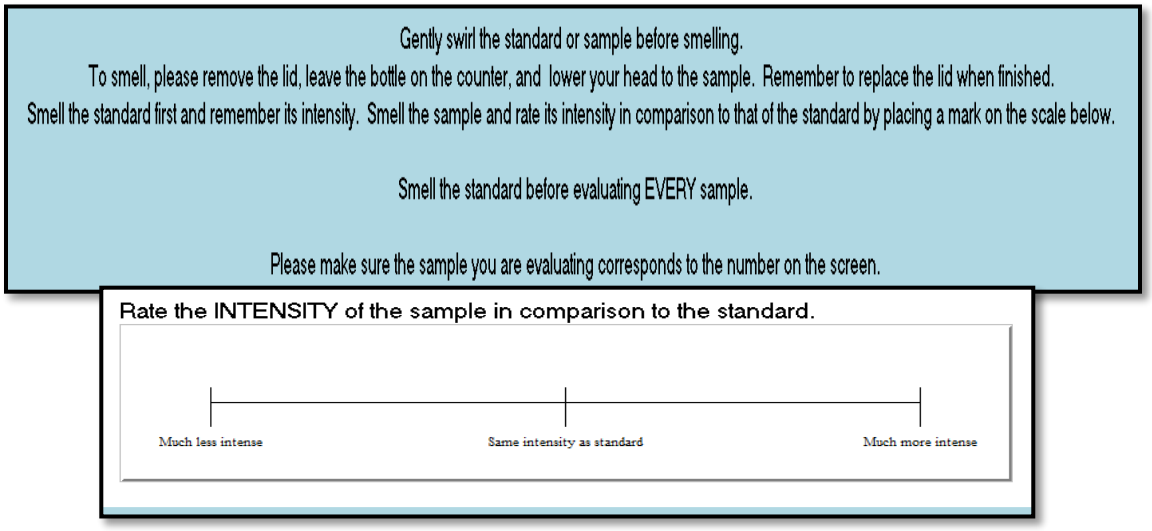


Figure 3.5: The SIMS Sensory Software instructions for intensity matching.

3.4 Data Analysis

SIMS Sensory Software provided sample intensity ratings for each panelist for butyric acid, furaneol, methional, and δ-decalactone. The ratings were a measure of the distance (in mm) from the left-most end on the line scale to the mark placed by panelists. We assigned the acetylpropionyl standard a rating of 75 on the intensity scale. This rating was the distance in mm from the "Much less intense" end of the scale to the "Same intensity as standard" mark in the center.

To obtain the concentrations at which all five compounds were at the same perceived intensity, we analyzed the data for each compound for each panelist individually. We used SAS© version 9.3 (Cary, NC) to perform regression analysis with concentration as the independent variable for butyric acid, furaneol, and δ-decalactone (Appendix G**:**). Because the concentrations for methional spanned several orders of magnitude, log concentration was used as the independent variable. Intensity rating was the dependent variable for all four compounds. We obtained a regression line as well as an equation for the line. We used Microsoft Excel® 2013 (Redmond, WA) to calculate the concentration matching the intensity of the standard by substituting 75, the value assigned to the intensity of the standard, for "y" in the equation and solving for "x." If it was necessary to extrapolate beyond the range of concentrations used for testing, we retested the panelist with the second set of concentrations (Table 3.2). If, after retesting, we still were not able to obtain a match, we determined the panelist was unable to

evaluate the odor compounds, and those panelists were not selected for the omission testing phase.

3.5 Results

We obtained the concentrations of butyric acid, furaneol, methional, and δ-decalactone matching the perceived intensity of 5 ppm acetylpropionyl for 22 panelists (Table 3.3). These panelists qualified for the omission-testing phase. The range of concentrations matching the intensity of the standard was: for butyric acid 58–188 ppm, for methional 0.003–0.396 ppm, for furaneol 100–514 ppm, and for δ-decalactone 28–202 ppm. Thirteen panelists were eliminated from the study because we were unable to obtain matches within the tested concentration range of at least one of the four compounds. The results for all panelists can be found in Appendix D. The concentrations of butyric acid, methional, furaneol, and δ-decalactone in Table 3.3 were used to prepare novel odor samples for each panelist in the second phase of the study, omission testing.

Table 3.3: Concentrations (ppm) of butyric acid, methional, furaneol, and δ-decalactone matching the intensity of 5 ppm acetylpropionyl for the 22 panelists who qualified for the omission testing phase.

Panelist	**Butyric Acid** (ppm)	**Methional** (ppm)	**Furaneol** (ppm)	**δ-Decalactone** (ppm)
1	133	0.276	103	49
2	82	0.003	169	59
3	107	0.651	175	58
4	101	0.010	225	68
5	93	0.074	216	30
6	103	0.003	239	126
7	103	0.044	167	74
8	188	0.045	192	93
9	78	0.004	114	38
10	109	0.049	254	83
11	141	0.242	137	28
12	129	0.192	325	103
13	100	0.101	328	44
14	107	0.180	472	114
15	101	0.073	167	119
16	159	0.396	318	168
17	80	0.049	178	112
18	181	0.008	514	202
19	59	0.074	100	37
20	58	0.117	100	188
21	188	0.101	129	37
22	78	0.199	136	71

4 Phase 2: Omission Testing

4.1 Participants

Of the 22 panelists who qualified to participate in omission testing, four chose not to participate. Eighteen panelists completed Phase 2.

4.2 Omission Samples

The concentrations obtained in Phase 1, intensity matching, provided a basis for the personalized testing samples we would need for our panel. We prepared 20 samples weekly for each panelist: 10 samples with the full mixture (all five compounds), five samples with one compound omitted (n-1 mixtures), and five with two compounds omitted (n-2 mixtures) (Table 4.1). We also prepared one additional full-mixture sample for each panelist labeled "Sample A."

Table 4.1: The 20 odor mixture samples presented to each panelist in each of the 20 omission testing sessions.

# of Samples Prepared	Omission Model	Sample Number	Sample Name	Omitted Compound(s)
10	n	S1-S10	n	none
5	n-1*	S11	n-BA	butyric acid
		S12	n-M	methional
		S13	n-F	furaneol
		S14	n-D	δ-decalactone
		S15	n-A	acetylpropionyl
5	n-2**	S16	n-BA,M	butyric acid methional
		S17	n-BA,A	butyric acid acetylpropionyl
		S18	n-M,D	methional δ-decalactone
		S19	n-D,F	δ-decalactone furaneol
		S20	n-F,A	furaneol acetylpropionyl

* Each compound was omitted from an n-1 sample once.
**The omissions for n-2 samples were chosen so each compound was eliminated an equal number of times.

The sample-mixing process began with the preparation of stock solutions of acetylpropionyl, butyric acid, furaneol, δ-decalactone, and methional. We prepared 250 mL of stock solution by pipetting the required amount of each compound into a volumetric flask and adding enough distilled water to total 250 mL (Table 4.2).

Table 4.2: Concentrations (ppm) of acetylpropionyl, butyric acid, furaneol, δ-decalactone, and methional stock solutions used for preparing the omission testing samples.

Compound	Concentration (ppm)	Concentration (mL/L)	Concentration (mL/250 mL)
Butyric Acid	4000	4	1.00
Methional	10	0.01	0.0025
Decalactone	4000	4	1.00
Acetylpropionyl	250	0.25	0.0625
Furaneol	8000	8 g	2.00 g

We used personalized sample-mixture plans created for each panelist to prepare individual samples from the stock solutions. The sample-mixture plans called for the required amounts of the necessary stock solutions combined with enough distilled water to create samples totaling 20 mL with appropriate compound concentrations (Figure 4.1). The plans were color-coded for each compound to facilitate sample preparation and reduce the chance of error.

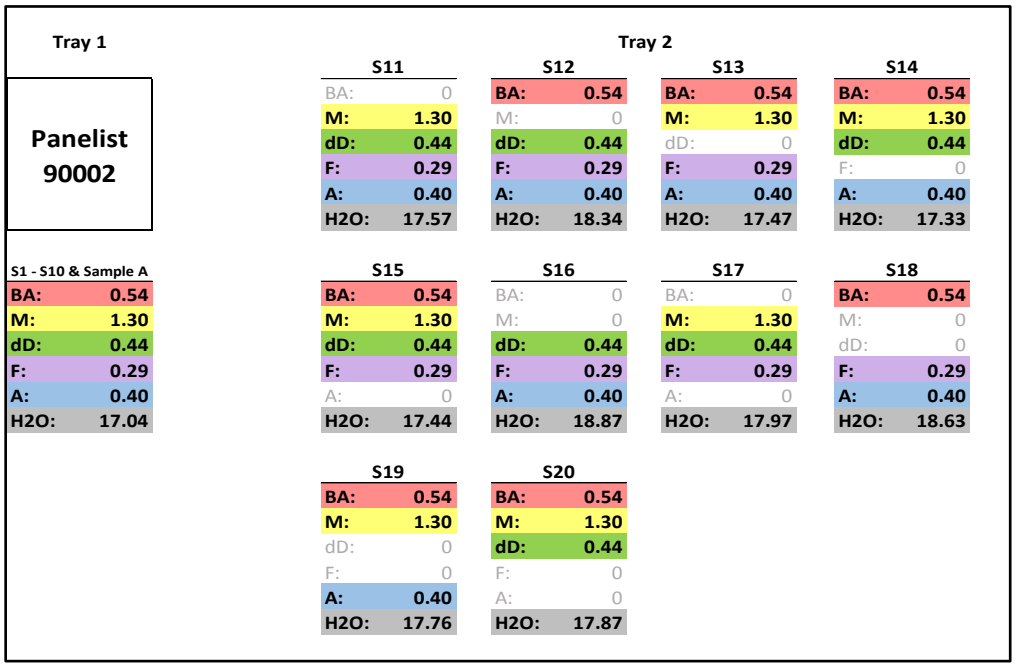


Figure 4.1: A personalized omission testing sample-mixture plan for one panelist. On the left is the amount of acetylpropionyl, butyric acid, furaneol, δ-decalactone, and methional stock solutions (mL) and distilled water required for the 10 full mixture samples and 'Sample A' for this panelist. The remainder of the plan gives the required amount of each stock solution and distilled water for the n-1 and n-2 samples.

We prepared the samples in 180 mL clear glass jars with unlined, polypropylene screw cap lids purchased from Qorpak®, A Division of Berlin Packaging (Bridgeville, PA). We coded each sample with random 3-digit numbers generated by SIMS Sensory Software. Each panelist was given a fresh preparation of samples weekly; but, within a given week, each panelist used the same set of samples for all sessions. We changed the codes after each session to prevent recognition of the samples (Appendix H).

We evaluated changes in the concentrations of the individual components of the mixtures over a one week period using Solid Phase Micro Extraction (SPME) followed by GC-MS to separate and identify the volatile compounds, and a Flame Ionization Detector (FID) to quantify (analysis was completed by Jean-Paul Schirle-Keller in the

Flavor Center at the University of Minnesota). Panelists evaluated samples both one and six days after preparation, so three one-day-old samples and three six-day-old samples were evaluated. A triphasic PDMS/DVB/Carboxen fiber (Supelco, Bellefonte, PA) was incubated at room temperature for one hour in a sealed 20 mL headspace vial containing x mL of solution. The resultant SPME isolates were analyzed on GC/MS 6890-5972 Agilent GC-MS. The SPME fibers were desorbed at 250°C for one minute into a GC/MS (6890-5972 Agilent GC-MS) operated in split (10:1) mode. Compounds were separated on a DB-Wax column (30 m long, 0.25 mm diameter, 0.25 µm film) using a temperature program beginning at 40 °C (2 min hold) and then increased to 230 °C (0 min) at 10 °C/min. The column effluent was split between the mass spectrometer and a PolyArc reactor interfaced with an FID detector. The mass spectrometer, PolyArc detector, and FID were at 230, 293 and 300°C respectively. The mass spectrometer was run in Scan mode to collect ions from 29 to 250 amu (3.34 scans/s). Mass spectra assured that we were measuring the correct model compounds.

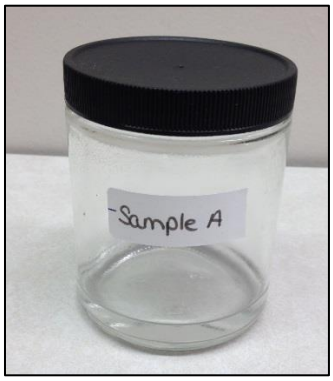


Figure 4.2: A prepared ‘Sample A’ for one panelist.

Prior to testing, we prepared trays for each panelist, labeled with the corresponding panelist's ID number and set up with panelist-specific 20 samples plus Sample A (Figure 4.3). We refrigerated the trays until approximately one hour before testing and then allowed the samples to come to room temperature.

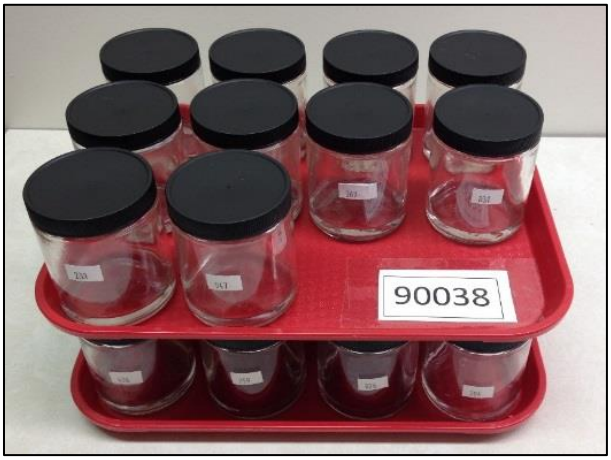


Figure 4.3: 20 omission testing samples prepared for one panelist.

4.3 Omission Testing Sessions

The panelists followed the same procedure for each of 20 sessions. Instructions and responses were communicated using SIMS Sensory Software on computers in sensory booths. The panelists were required to log into the proper test for each session on the booth computers. To ensure that they were properly logged in, we prepared Panelist ID Cards containing their ID numbers (used for logging into the software and tracking responses) and session numbers. Each card was divided into sections, and each section contained the necessary information for each of the 20 sessions. When panelists arrived for a session, we simply removed the corresponding section of the cards for the

current session and gave it to them. They could then use the information to begin the test session (Figure 4.4).

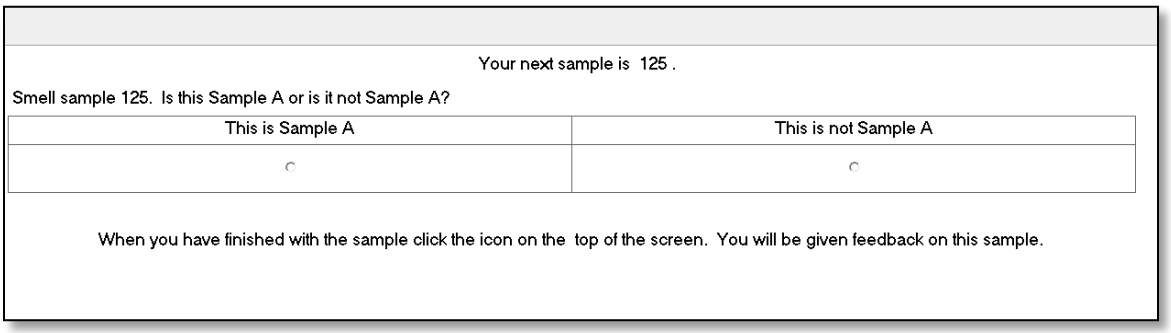
Panelist 339 **Session 1**	**Panelist 339** **Session 11**
Panelist 339 **Session 2**	**Panelist 339** **Session 12**
Panelist 339 **Session 3**	**Panelist 339** **Session 13**
Panelist 339 **Session 4**	**Panelist 339** **Session 14**
Panelist 339 **Session 5**	**Panelist 339** **Session 15**
Panelist 339 **Session 6**	**Panelist 339** **Session 16**
Panelist 339 **Session 7**	**Panelist 339** **Session 17**
Panelist 339 **Session 8**	**Panelist 339** **Session 18**
Panelist 339 **Session 9**	**Panelist 339** **Session 19**
Panelist 339 **Session 10**	**Panelist 339** **Session 20**

Figure 4.4: Omission testing panelist ID cards.

To evaluate the samples, panelists participated in a series of A-not-A Tests with feedback. We instructed the panelists to first smell "Sample A" and remember the odor. They were then instructed to smell the first test sample and determine if it was "Sample A" or "Not Sample A." The panelists responded by selecting their choices on the screen. The panelists were given feedback via the computer immediately after responding to each

sample. They were told that the sample "was Sample A" or "was Not Sample A" (Figure 4.5). The panelists were then required to wait 30 seconds before evaluating the next sample to minimize adaptation. The sample order was balanced among panelists using a Latin square design (Williams, 1949) for 20 samples. Both the wait time and the sample order were enforced by the software. The panelists followed this procedure for all 20 samples. Each session lasted approximately 30 minutes.

a)



b)

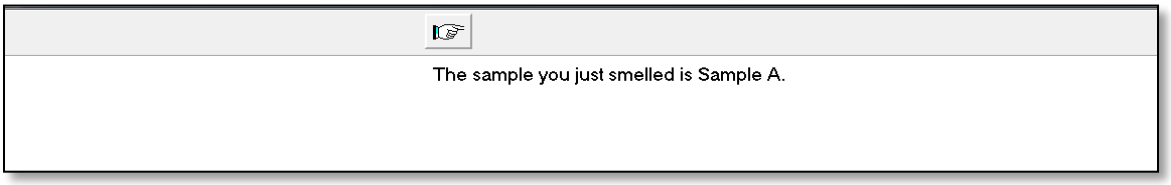


Figure 4.5: a) An example of the A-not-A Test directions given by SIMS Sensory Software during omission testing. **b)** An example of the feedback given following the panelists' response.

4.4 Data Analysis

We used signal detection theory and the discriminability index (d') to evaluate the omission testing data by sessions, panelists, and omitted compounds. Because the

samples of interest in this study were the omission samples (n-1 or n-2), a hit was defined as responding "Not A" when an omission sample was presented. A false alarm was defined as responding "Not A" when a full mixture sample was presented.

4.4.1 Session

We analyzed the data by sessions to determine whether discrimination improved as the sessions progressed. Our thought was that if the d' values for each session increased as the sessions progressed, learning had occurred. We pooled the data for all panelists within each session. We calculated d' values for each session using R 3.3.1 (R Core Team, 2016) using the car (Fox & Weisberg, 2011), geepack (Hojsgaard, Halekoh, & Yan, 2006), and xlsx (Dragulescu, 2014) packages. We used Equations (4.1 and (4.2. to calculate the proportion of n-1 hits, n-2 hits, and false alarms.

$$proportion_{hits} = \frac{\#\ of\ hits}{\#\ of\ possible\ hits} \tag{4.1}$$

$$proportion_{false\ alarms} = \frac{\#\ of\ false\ alarms}{\#\ of\ possible\ false\ alarms} \tag{4.2}$$

From these proportions we obtained the corresponding z-values. We calculated the discriminability index (d') for the n-1 and n-2 mixtures for each session using Equation (4.3.

$$d' = z_{hits} - z_{false\ alarms} \tag{4.3}$$

A d' value of zero would indicate that the odor samples were perceptually equal and panelists were not able to distinguish between them. Because the observations were repeated binary observations, we used a Generalized Estimating Equations (GEE) model with an exchangeable correlation structure for each session/judge combination. A probit link function allowed us to directly estimate d'. We used Wald 95% confidence intervals to determine if the d' values we obtained were statistically significantly different from zero. We completed the analysis using R 3.3.1 (Appendix J).

4.4.2 Panelist

We analyzed the data for each panelist to determine if that panelist could discriminate between full-mixture samples and samples with omitted compounds. We analyzed data for the n-1 and n-2 mixtures separately. A review of the results for sessions shows that the shape of the trend line changed after session five and leveled out between sessions five and 10. Therefore we determined that after 10 sessions they had enough training to become familiarized with the mixtures; and, because we were interested in whether they could discriminate, we used data from only these last 10 sessions.

We analyzed the data for panelists in a manner similar to that for sessions (see section 4.4.1 above); we calculated proportions of n-1 hits, n-2 hits, and false alarms for each panelist, obtained z-scores, and calculated d'. The d' values were analyzed

statistically using a GEE model and Wald 95% confidence intervals. We completed the analysis using R 3.3.1 (Appendix J).

4.4.3 Compound

We analyzed the data by omitted compound to determine if discrimination was compound dependent. A review of the results for sessions (Figure 4.6) shows that the shape of the trend line changed after session five and leveled out between sessions five and 10. Therefore, we determined that after 10 sessions the panelists had enough training to become familiarized with the mixtures; and, because we were interested in whether they could discriminate, we used data from only these last 10 sessions.

We analyzed the data for each compound omitted in a similar manner as above (see section 4.4.1) with one exception. Because analysis of the data we obtained from n-1 omission samples would indicate whether discrimination was compound dependent, we did not analyze the data for n-2 samples. We calculated proportions of n-1 hits and false alarms for each omitted compound, obtained z-scores, and calculated d'. The d' values were analyzed statistically using a GEE model and Wald 95% confidence intervals. To determine whether there was a difference between compounds we analyzed the results using a Tukey HSD multiple comparison. We completed the analysis using R 3.3.1 (Appendix J).

4.5 Results

4.5.1 Session

The discrimination index (d') estimated values for n-1 omission samples obtained for each session ranged from 0.092 (Session 1) to 0.897 (Session 9). The d' values for 15 of the 20 sessions were significantly greater than zero ($p < 0.05$). The d' values for each session, as well as statistical analysis results, can be found in Table 4.3.

The d' estimated values obtained for n-2 omission samples, ranged from 0.717 (Session 1) to 1.368 (Session 7). These values were significantly different from zero ($p < 0.05$) for all 20 sessions. The results for each session can be found in Table 4.4.

The d' values for n-2 omission samples were greater than those for n-1 samples for all sessions. The confidence intervals for n-1 samples and n-2 samples had very little or no overlap in every session. A comparison of d' values for both n-1 and n-2 omission samples can be found in Figure 4.6**:**.

Table 4.3: Discrimination index (d') estimated values for n-1 omission samples for each session. We calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. We used Wald 95% confidence intervals to determine if the d' values we obtained were statistically significantly different from zero. The squared Wald statistic and corresponding p-value, are reported here as well as the upper and lower confidence intervals.

Session	d'	SE	Wald*	p-value	lower	upper
1	0.09	0.169	0.30	0.584	-0.24	0.42
2	0.50**	0.169	8.79	0.003	0.17	0.83
3	0.20	0.112	3.21	0.073	-0.02	0.42
4	0.50**	0.172	8.44	0.004	0.16	0.84
5	0.58**	0.127	20.67	< 0.001	0.33	0.83
6	0.37**	0.148	6.26	0.012	0.08	0.66
7	0.58**	0.142	17.00	< 0.001	0.31	0.86
8	0.32**	0.144	4.98	0.026	0.04	0.60
9	0.90**	0.125	51.28	< 0.001	0.65	1.14
10	0.31	0.176	3.14	0.076	-0.03	0.66
11	0.83**	0.233	12.65	< 0.001	0.37	1.29
12	0.57**	0.189	9.17	0.002	0.20	0.94
13	0.36	0.198	3.36	0.067	-0.03	0.75
14	0.56**	0.161	12.27	< 0.001	0.25	0.88
15	0.60**	0.190	10.03	0.002	0.23	0.97
16	0.63**	0.197	10.12	0.001	0.24	1.01
17	0.42**	0.182	5.26	0.022	0.06	0.77
18	0.21	0.212	0.94	0.332	-0.21	0.62
19	0.44**	0.164	7.16	0.007	0.12	0.76
20	0.38**	0.183	4.26	0.039	0.02	0.74

*The Wald statistics have been squared to remove negative values.
**Indicates a d' value that is statistically significantly different from 0 (p-value < 0.05).

Table 4.4: Discrimination index (d') estimated values for n-2 omission samples for each session. We calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. We used Wald 95% confidence intervals to determine if the d' values we obtained were statistically significantly different from zero. The squared Wald statistic and corresponding p-value, are reported here as well as the upper and lower confidence intervals.

Session	d'	SE	Wald*	p-value	lower	upper
1	0.72**	0.171	17.52	< 0.001	0.38	1.05
2	0.78**	0.172	20.74	< 0.001	0.45	1.12
3	1.05**	0.178	35.09	< 0.001	0.70	1.40
4	1.05**	0.222	22.36	< 0.001	0.61	1.48
5	1.13**	0.188	36.28	< 0.001	0.76	1.50
6	1.14**	0.162	49.80	< 0.001	0.82	1.46
7	1.37**	0.185	54.61	< 0.001	1.01	1.73
8	1.06**	0.163	42.39	< 0.001	0.74	1.38
9	1.27**	0.199	40.88	< 0.001	0.88	1.66
10	0.99**	0.186	28.61	< 0.001	0.63	1.36
11	1.12**	0.281	15.83	< 0.001	0.57	1.67
12	1.07**	0.185	33.36	< 0.001	0.70	1.43
13	1.10**	0.210	27.48	< 0.001	0.69	1.52
14	1.48**	0.159	86.71	< 0.001	1.17	1.79
15	1.25**	0.207	36.31	< 0.001	0.84	1.65
16	1.30**	0.235	30.55	< 0.001	0.84	1.76
17	1.11**	0.150	55.08	< 0.001	0.82	1.41
18	1.07**	0.208	26.30	< 0.001	0.66	1.48
19	1.19**	0.215	30.49	< 0.001	0.76	1.61
20	1.10**	0.221	24.57	< 0.001	0.66	1.53

*The Wald statistics have been squared to remove negative values.
**Indicates a d' value that is statistically significantly different from 0 (p-value < 0.05).

Using a logarithmic trend line for best fit with the data for both n-1 omission samples and n-2 omission samples, learning occurred during the first few sessions as indicated by the increasing slope of the trend line (Figure 4.6). By Session four the slope of the line began to decrease.

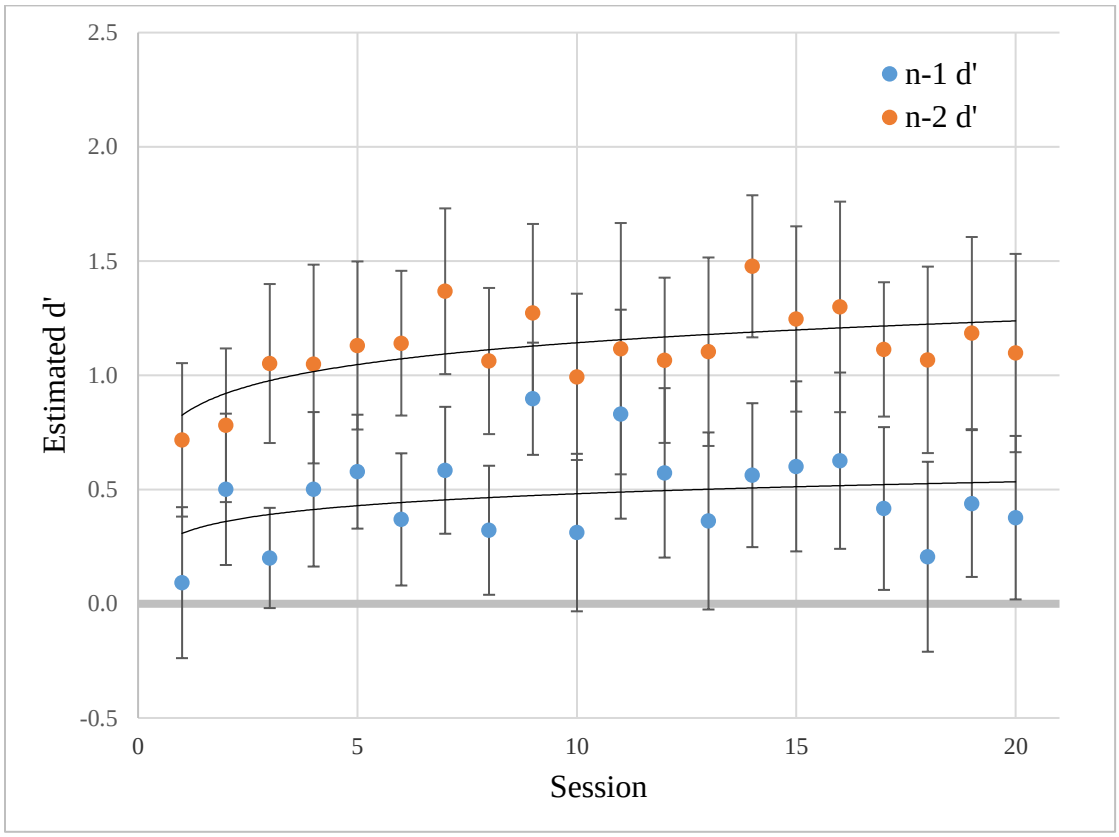


Figure 4.6: Discrimination index (d') estimated values for n-1 and n-2 omission samples for each of 20 sessions. Each panelist (n=18) evaluated 20 samples per session. We pooled the data for all panelists within each session, and we calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. Error bars are Wald 95% confidence intervals.

4.5.2 Panelist

The estimated discrimination index (d') values we obtained for the last 10 sessions for each panelist ranged from -0.230 to 1.077 (n-1 omission samples) and 0.131 to 2.196 (n-2 omission samples). The d' values were statistically significantly different from zero for 11 of the 18 panelists for n-1 omission samples and for 17 of the 18

panelists for n-2 samples. The results for each panelist for n-1 and n-2 omission samples can be found in Table 4.5 and Table 4.6 respectively.

Table 4.5: Discrimination index (d') estimated values for n-1 omission samples for each panelist. We calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. We used Wald 95% confidence intervals to determine if the d' values we obtained were statistically significantly different from zero. The squared Wald statistic and corresponding p-value, are reported here as well as the upper and lower confidence intervals.

Panelist	d'	SE	Wald*	p-value	lower	upper
1	0.51**	0.222	5.23	0.022	0.07	0.94
2	0.44	0.245	3.29	0.070	-0.04	0.93
3	0.37	0.252	2.15	0.142	-0.12	0.87
4	0.18	0.272	0.45	0.501	-0.35	0.72
5	0.63**	0.255	6.02	0.014	0.13	1.12
6	0.83**	0.244	11.61	0.001	0.35	1.31
7	0.59**	0.152	14.90	< 0.001	0.29	0.88
8	-0.23	0.177	1.69	0.194	-0.58	0.12
9	0.51**	0.220	5.43	0.020	0.08	0.94
10	0.63**	0.221	8.08	0.004	0.20	1.06
11	0.85**	0.269	9.89	0.002	0.32	1.38
12	0.98**	0.219	20.12	< 0.001	0.55	1.41
13	1.08**	0.236	20.80	< 0.001	0.61	1.54
14	0.39**	0.167	5.48	0.019	0.06	0.72
15	0.31	0.253	1.46	0.228	-0.19	0.80
16	0.61**	0.259	5.53	0.019	0.10	1.12
17	0.03	0.253	0.02	0.902	-0.46	0.53
18	0.15	0.216	0.50	0.480	-0.27	0.58

*The Wald statistics have been squared to remove negative values.
**Indicates a d' value that is statistically significantly different from 0 (p-value < 0.05).

Table 4.6: Discrimination index (d') estimated values for n-2 omission samples for each panelist. We calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. We used Wald 95% confidence intervals to determine if the d' values we obtained were statistically significantly different from zero. The squared Wald statistic and corresponding p-value, are reported here as well as the upper and lower confidence intervals.

Panelist	d'	SE	Wald*	p-value	lower	upper
1	1.49**	0.196	57.94	< 0.001	1.11	1.87
2	1.55**	0.271	32.70	< 0.001	1.02	2.08
3	1.31**	0.252	27.16	< 0.001	0.82	1.81
4	1.18**	0.240	24.23	< 0.001	0.71	1.65
5	1.37**	0.311	19.21	< 0.001	0.75	1.98
6	1.72**	0.229	56.58	< 0.001	1.27	2.17
7	1.08**	0.235	21.18	< 0.001	0.62	1.55
8	0.49**	0.182	7.23	0.007	0.13	0.85
9	1.18**	0.201	34.33	< 0.001	0.78	1.57
10	0.61**	0.204	8.81	0.003	0.21	1.01
11	1.52**	0.213	50.66	< 0.001	1.10	1.93
12	2.20**	0.303	52.49	< 0.001	1.60	2.79
13	2.07**	0.239	75.06	< 0.001	1.60	2.54
14	0.87**	0.177	23.81	< 0.001	0.52	1.21
15	1.01**	0.261	14.94	< 0.001	0.50	1.52
16	0.75**	0.255	8.62	0.003	0.25	1.25
17	0.56**	0.207	7.23	0.007	0.15	0.96
18	0.13	0.205	0.41	0.523	-0.27	0.53

*The Wald statistics have been squared to remove negative values.
**Indicates a d' value that is statistically significantly different from 0 (p-value < 0.05).

A comparison of panelists' d' values for n-1 and n-2 samples (Figure 4.7) indicates that seven panelists had d' values that were significantly higher when two compounds were omitted from the sample (n-2) than one (n-1) as indicated by confidence intervals that were not overlapping. (Figure 4.7.) The d' values for n-1 and n-2 samples were not significantly different for the remaining panelists as indicated by overlapping confidence intervals.

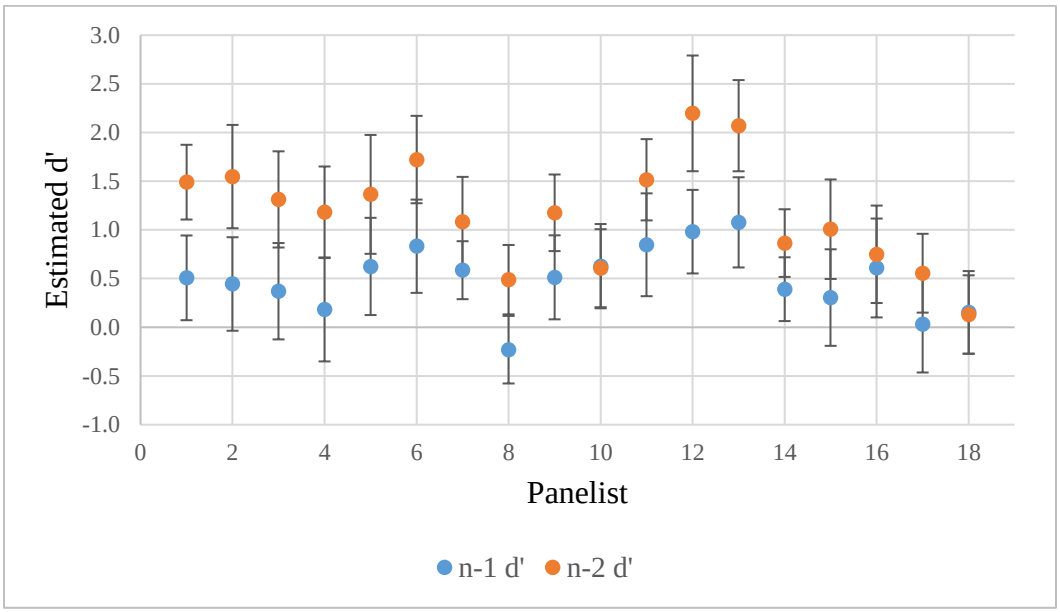


Figure 4.7: Discrimination index (d') estimated values for n-1 and n-2 omission samples for each panelist. Each panelist evaluated 20 samples in each of 20 sessions. We pooled the data for the last 10 sessions, and we calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. Error bars are Wald 95% confidence intervals.

4.5.3 Compound

The estimated d' values for omitted compounds, obtained by pooling data for n-1 samples from the last 10 sessions, ranged from 0.367 (furaneol) to 0.695 (δ-decalactone). The d' values were all statistically significantly different from zero. The analysis results can be found in Table 4.7 and are shown in Figure 4.8. The results of the Tukey HSD test indicate that the d' values for all five omitted compounds were not significantly different (Table 4.8).

Table 4.7: Discrimination index (d') estimated values for n-1 omission samples. We calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. We used Wald 95% confidence intervals to determine if the d' values we obtained were statistically significantly different from zero. The squared Wald statistic and corresponding p-value, are reported here as well as the upper and lower confidence intervals.

Compound	d'	SE	Wald*	p-value	lower	upper
Butyric Acid	0.43**	0.102	18.05	< 0.001	0.23	0.63
δ-decalactone	0.70**	0.102	46.17	< 0.001	0.50	0.90
Acetylpropionyl	0.52**	0.099	27.18	< 0.001	0.32	0.71
Furaneol	0.37**	0.103	12.69	< 0.001	0.17	0.57
Methional	0.40**	0.097	17.22	< 0.001	0.21	0.60

*The Wald statistics have been squared to remove negative values.
**Indicates a d' value that is statistically significantly different from 0 (p-value < 0.05).

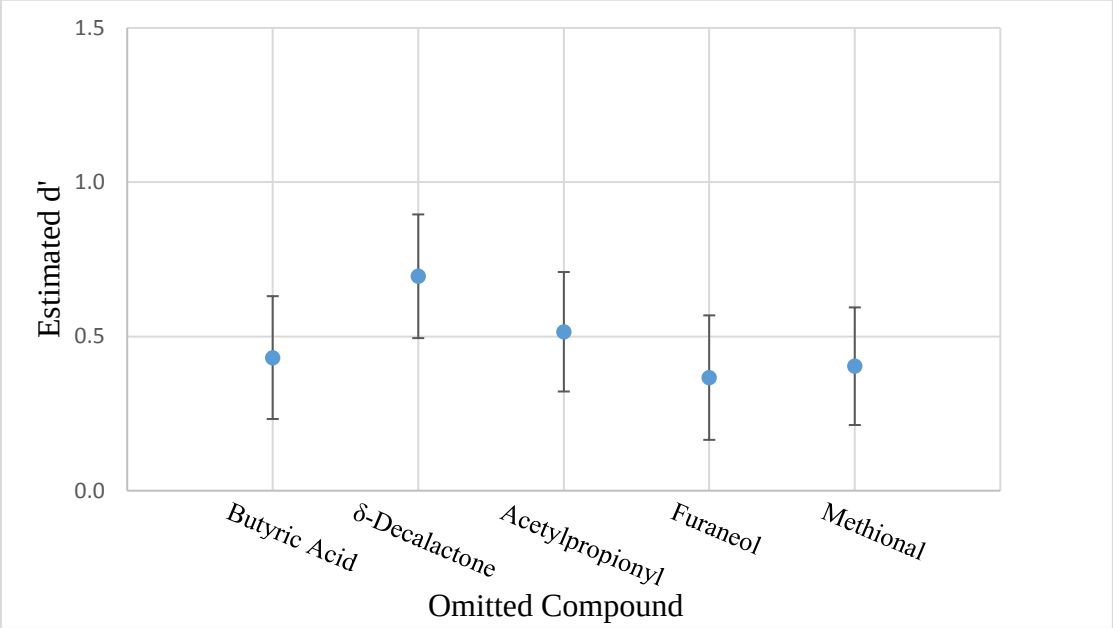


Figure 4.8: Discrimination index (d') estimated values for n-1 omission samples for each omitted compound. Each panelist (n=18) evaluated 20 samples per session for 20 sessions. We pooled the data for all panelists for the last 10 sessions, and we calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. Error bars are Wald 95% confidence intervals.

Table 4.8: Results of the Tukey HSD test comparing differences in estimated d' values between pairs of omitted compounds. All paired combinations of acetylpropionyl (A), butyric acid (BA), furaneol (F), δ-decalactone (DD), and methional (M) were tested.

Tukey HSD Tests	p-value
BA - DD	0.334
BA -A	0.976
BA - F	0.991
BA - M	1.000
DD - A	0.697
DD - F	0.135
DD - M	0.213
A - F	0.829
A - M	0.928
F - M	0.999

Instrumental analysis of three six-day-old samples and three one-day-old samples using SPME followed by GC-MS indicated that there was a significant decrease in the concentration of methional but not acetylpropionyl, butyric acid, furaneol, or δ-decalactone (Figure 4.9).

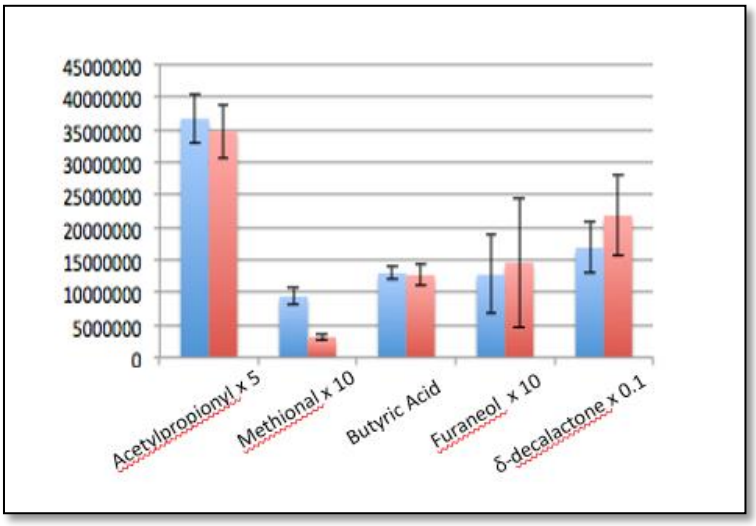


Figure 4.9: The mean results of the instrumental analysis of three one-day-old (on left) and three six-day-old (on right) omission testing samples. We evaluated changes in the individual component concentrations of each sample using Solid Phase Micro Extraction (SPME) followed by GC-MS to separate and identify the volatile compounds, and a Flame Ionization Detector (FID) to quantify. The x-axis indicates the omitted compound (scores were multiplied by a factor of 0.1, 5, or 10 to place them on the same scale) and the y-axis indicates peak area. Error bars are the standard deviation of the sample replicates.

5 Discussion

In order to discriminate between a given odor mixture and others with similar olfactory receptor neuron (ORN) patterns, training is necessary. Stevenson and associates (Stevenson & Mahmut, 2013; Stevenson & Wilson, 2007b) demonstrated that familiar patterns of ORN activity, which are stored in memory, are activated during discrimination tasks. Their research suggests that successful discrimination depends on matching incoming patterns with those stored in memory. Further, according to Soto & Ashby's (2015) research on category learning, the process of storing odor activity patterns in memory is implicit. One strategy that has been shown to help such implicit

learning involves training panelists by exposing them to the odor mixtures of interest and providing immediate feedback on the correctness of their responses (Ell et al., 2012). The panelists in our study had no prior exposure to the full five-compound mixture or any of the mixtures with one of the five compounds removed. Therefore, they had no matching patterns stored in memory. Thus, the panelists were not able to discriminate between the mixtures initially. However, following repeated exposure to the mixtures and feedback on the correctness of their answers, the panelists collectively were able to discriminate between the mixtures indicating that patterns of receptor stimuli had been stored in memory. The panelists' ability to learn was evidenced by an initial increase in the slope of the logarithmic trend line for n-1 omission samples for the first few sessions (Figure 4.6). A flattening of the trend line indicated that about four sessions of 20 trials were needed to maximize learning.

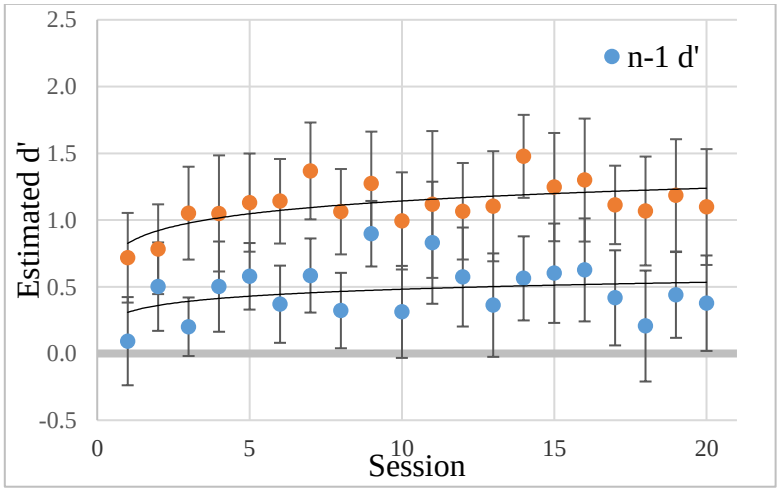


Figure 5.1: Discrimination index (d') estimated values for n-1 and n-2 omission samples for each of 20 sessions. Each panelist (n=18) evaluated 20 samples per session. We pooled the data for all panelists within each session, and we calculated the estimated d' values and standard errors using a Generalized Estimating Equations model. Error bars are Wald 95% confidence intervals.

We expected that pattern completion during processing would inhibit the panelists' ability to discriminate between the full five-compound mixture and the same mixture with one missing compound. Odor mixtures containing compounds with similar perceived intensities, such as those used in this study, are processed as unique odor objects, according to Livermore and Laing (1998b) and Sinding and associates (2013b). Wilson and Sullivan's study (2011) suggested that when panelists encounter familiar odor mixtures in which one compound has been omitted, their brain recognizes the pattern of stimuli for the full mixture. That work suggests that their brain completes the pattern by filling in the missing stimuli. This pattern completion would prevent panelists from discriminating between the full mixture and mixtures with one missing compound. Low d' values indicated that discrimination was difficult. What's more, seven of the 18 panelists who completed omission testing were unable to discriminate the mixtures from one another.

The more compounds present in an odor mixture, the less likely people should be to discriminate a partial mixture from the full mixture. When Jinks and Laing (2001) increased the number of compounds in their study's odor mixtures from one to four, panelists no longer could identify individual compounds. The researchers suggested that increasing the number of compounds in a mixture increases the complexity of the pattern of stimuli processed in the piriform cortex. Rats in a study completed by Barnes and associates (2008) were unable to discriminate between mixtures with 10 compounds and those with one of the 10 removed. Removing one compound from a mixture of 10 would result in the loss of 10% of the signal being sent to the cortical region of the brain. The

researchers concluded that the rats' brains had completed the pattern of receptor activity, filling in the gaps for the missing signals and preventing discrimination. Removing one compound from a mixture of five, as was done in this study, would result in the loss of 20% of the signal to the piriform cortex. A 20% loss may result in a pattern that is different enough from the original to prevent pattern completion from occurring and allow for discrimination. However, the majority of recombination models used for omission testing in the studies we reviewed combined between 19 and 53 compounds (Burdack-Freitag & Schieberle, 2012; Czerny et al., 1999; Escudero et al., 2004; Ferreira et al., 2002; Gao et al., 2014; Kiatbenjakul et al., 2015; Mayer et al., 2000; Pavez et al., 2015). Omitting one compound from 53 would only involve a loss of 2% of the stimulus, allowing for pattern completion to occur when compounds that have blended or fused are removed.

The results we obtained in this study are consistent with those obtained from omission testing done by Hongsoongnern (2003). We used the raw data obtained by Hongsoongnern to calculate estimated d' values for each session and for each panelist for the last 10 sessions. The ranges of estimated d' values for each session and each panelist were similar (Table 5.1).

Table 5.1: Comparison of the range of estimated d' values for sessions and panelists in the omission testing study completed by Hongsoongnern (2003) (n=9) and this study (2016) (n=18). Estimated d' values for the 2003 study were calculated from raw data obtained by Hongsoongnern. Estimated d' values for each panelist were calculated using data from the last 10 sessions for both studies.

	Estimated d'	
	2003 Study	2016 Study
Sessions	0.15 – 1.05	0.09 – 0.90
Panelists	0.36 – 1.11	-0.23 – 1.08

Hongsoongnern calculated the average total number of correct responses for each session rather than estimating d'. However, a comparison of the plotted results for each session shows a similar curve in the logarithmic trendlines for both studies (Figure 5.2). Both studies yielded an initial upward slope in the trend line indicating that panelists' performance was improving. This initial upward slope was followed by a flattening of the line indicating that performance no longer was improving.

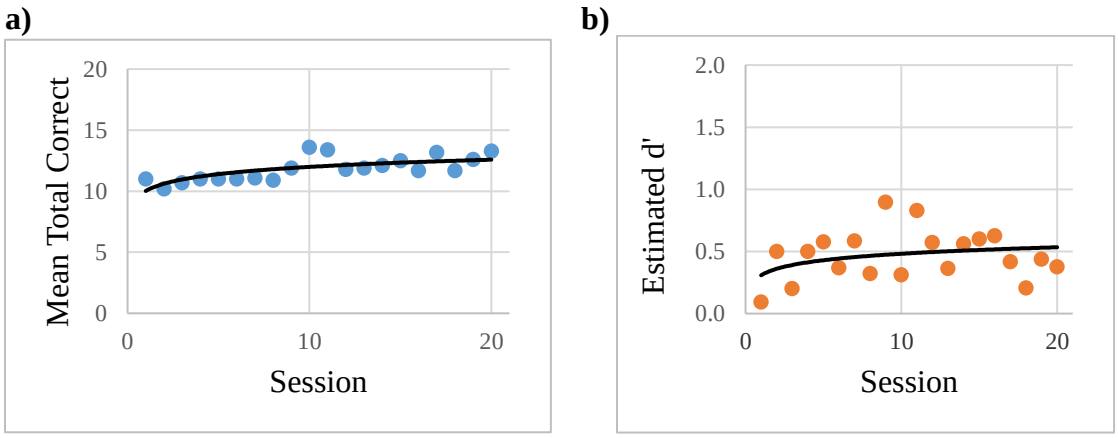


Figure 5.2: a) A plot of the mean total correct responses for each session from the omission testing study completed by Hongsoongnern (2003) (N = 9). Participants evaluated 20 samples in each session. Ten samples were the full five-compound mixtures, and 10 were the same mixture with one of the five compounds omitted. Each compound was omitted from two samples. **b)** A plot of estimated d' values for n-1 omission samples for each session from this study (N = 18). The d' values were calculated from pooled data from the evaluation of 10 full five-compound mixture samples and five n-1 mixtures.

The low levels of d' found to be statistically significant in this study indicated that an experimental design with high statistical power may be required to detect some differences between samples in omission testing. Many of the omission testing studies we reviewed likely missed finding some differences for omissions when actual

differences existed because the studies had low statistical power. A discrimination test, such as a triangle test, can be designed to minimize the risk of missing a true difference by calculating the appropriate sample size required for a chosen levels of power and level of significance. Calculating the appropriate sample size requires the probability of a correct guess (P_{cg}) for the difference test, and the proportion of the panelists who are able to discriminate (P_d) as well as the desired level of power and sensitivity. In this study we found that panelists could discriminate between the full mixture and partial mixtures with any one compound removed, but the estimated d' values we obtained for each session were very low. The lowest significant d' value we obtained was 0.32. We can calculate the sample size (or number of panelists) that would be required to detect this difference using triangle testing, the method most commonly used in the omission studies we reviewed. A study completed by Ennis (1993) compared the power of various discrimination tests, including the triangle test, at different values of d'. According to Ennis, the probability of a correct response (P_c) using triangle testing at a d' equal to 0.32 would be 0.3427 (Ennis, 1993). The probability of a correct response (P_c) also can be calculated using Equation (5.1,

$$P_c = (P_d * P_{cd}) + (P_g * P_{cg}) \quad (5.1)$$

where P_d is the proportion of panelists who can discriminate between samples, P_{cd} is the probability of a correct response from a panelist who can discriminate between the samples (P_{cd} always equals 1.0), P_g is the proportion of panelists who are guessing, and P_{cg} is the chance probability of a correct guess. In difference testing, correct responses

occur either because panelists can discriminate between the samples or because they guess correctly. The proportion of panelists who are discriminators (P_d) plus the proportion who are guessing (P_g) must total one Equation (5.2.

$$P_d + P_g = 1 \tag{5.2}$$

The proportion of panelists guessing correctly can be calculated using Equation (5.3.

$$P_g = 1 - P_d \tag{5.3}$$

We can substitute the proportion of panelists guessing correctly (P_g) in Equation (5.1with Equation (5.3to get Equation (5.4.

$$P_c = (P_d * 1.0) + ((1 - P_d) * P_{cg}) \tag{5.4}$$

Solving for the proportion of discriminators (P_d) we get Equation (5.5.

$$P_d = (P_c - P_{cg})/(P_{cd} - P_{cg}) \tag{5.5}$$

In a triangle test, the chance of guessing correctly (P_{cg}) is 1/3 or 0.33. The probability of a correct response for discriminators (P_{cd}) is always 1.0. Using Equation 6.5, we find that the probability of a correct response (P_c) equal to 0.3427 would result in a proportion of distinguishers (P_d) equal to 0.018. The sample size (or number of panelists) required to find a difference similar to the one found in this study using triangle tests with the chance of guessing correctly (P_{cg}) equal to 0.33 and a proportion of distinguishers (P_d) equal to 0.018 would be 7664 (power = 0.80, $\alpha = 0.05$) (Meilgaard et al., 2016). Even then, at a power of 0.80, there would be a 20% chance that the

difference would go undetected. Increasing the power to 0.95 would result in a required sample size of 13458. The probability of a correct response (P_c) using triangle testing at a d' equal to 1.0 would be 0.4180 (Ennis, 1993), the proportion of distinguishers (P_d) would be 0.13 (Equation (5.5), and the number of panelists required would be 187 (power = 0.80, $\alpha = 0.05$) or 331 (power = 0.95, $\alpha = 0.05$) (Meilgaard et al., 2016).

Changes in the concentration of methional over time did not seem to impact the results of this study. The results we obtained from analysis of samples of the full five-compound mixture using SPME followed by GC-MS indicated that there was a significant decrease in methional concentration over a five day period. A diminishing concentration of methional in the full mixture samples as the week progressed would result in an increasing similarity between the full mixtures and the partial mixtures in which methional was omitted. However, the results of Tukey's HSD test indicated that the estimated d' values calculated for each compound did not differ from one another. Panelists' performance was the same regardless of which compound was omitted.

Omission testing as a method for determining the important odor compounds in a mixture can be an effective method but should be used with caution. In this study, the panelists were able to discriminate between the full five-compound mixture and the same mixture with any one of the five compounds omitted. However, the panelists were screened for their ability to evaluate the individual compounds, and they underwent substantial training before omission testing results were evaluated. The screening methods and extensive training would increase the proportion of participants who were discriminators, thus increasing the power of the test. Even with a statistically powerful

test learning was not impressive. Estimated d' values were very low indicating that the samples were perceptually very similar and discrimination was difficult. When compounds are at a similar perceived intensity, using omission testing, as a method for determining the important odorants in a mixture, with too few panelists who are untrained and have not been screened, is likely to miss finding a difference when a difference exists.

6 Conclusion

The statistical results indicated that the panelists, as a group, were able to discriminate between the full five-compound mixture and mixtures with any one of the five compounds removed. However, low d' values indicated that the mixtures were extremely difficult to discriminate. What's more, 39% of the panelists, who were preselected for their ability to evaluate these compounds, never could discriminate between the full mixture and partial mixtures.

References

Abdi, H. (2007). *Signal Detection Theory (SDT)*.

Anderson, N. D. (2015). Teaching signal detection theory with pseudoscience. *Frontiers in Psychology*, *6*, 762. http://doi.org/10.3389/fpsyg.2015.00762

Atanasova, B., Thomas-Danguin, T., Chabanet, C., Langlois, D., Nicklaus, S., & Etiévant, P. (2005). Perceptual interactions in odour mixtures: odour quality in binary mixtures of woody and fruity wine odorants. *Chemical Senses*, *30*(3), 209–17. http://doi.org/10.1093/chemse/bji016

Barkat, S., Le Berre, E., Coureaud, G., Sicard, G., & Thomas-Danguin, T. (2012). Perceptual blending in odor mixtures depends on the nature of odorants and human olfactory expertise. *Chemical Senses*, *37*(2), 159–66. http://doi.org/10.1093/chemse/bjr086

Barnes, D. C., Hofacer, R. D., Zaman, A. R., Rennaker, R. L., & Wilson, D. A. (2008). Olfactory perceptual stability and discrimination. *Nature Neuroscience*, *11*(12), 1378–80. http://doi.org/10.1038/nn.2217

Benkwitz, F., Nicolau, L., Lund, C., Beresford, M., Wohlers, M., & Kilmartin, P. A. (2012). Evaluation of key odorants in sauvignon blanc wines using three different methodologies. *Journal of Agricultural and Food Chemistry*, *60*(25), 6293–302. http://doi.org/10.1021/jf300914n

Buck, L. B. (2000). The Molecular Architecture of Odor and Pheromone Sensing in Mammals. *Cell*, *100*(6), 611–618. http://doi.org/10.1016/S0092-8674(00)80698-4

Buck, L. B. (2004). Olfactory Receptors and Odor Coding in Mammals. *Nutrition Reviews*, *62*(s3), S184–S188. http://doi.org/10.1111/j.1753-4887.2004.tb00097.x

Buck, L. B. (2013). Smell and Taste: The Chemical Senses. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, & A. J. Hudspeth (Eds.), *Principles of Neural Science* (5th ed., pp. 712–742). McGraw-Hill Companies, Inc.

Burdack-Freitag, A., & Schieberle, P. (2010). Changes in the key odorants of Italian Hazelnuts (Coryllus avellana L. Var. Tonda Romana) induced by roasting. *Journal of Agricultural and Food Chemistry*, *58*(10), 6351–9. http://doi.org/10.1021/jf100692k

Burdack-Freitag, A., & Schieberle, P. (2012). Characterization of the key odorants in raw Italian hazelnuts (Corylus avellana L. var. Tonda Romana) and roasted hazelnut paste by means of molecular sensory science. *Journal of Agricultural and Food Chemistry*, *60*(20), 5057–64. http://doi.org/10.1021/jf300908d

Czerny, M., Mayer, F., & Grosch, W. (1999). Sensory Study on the Character Impact Odorants of Roasted Arabica Coffee. *Journal of Agricultural and Food Chemistry*, *47*(2), 695–699. http://doi.org/10.1021/jf980759i

Delime, P., O'Mahony, K., Desforges, N., Taylor, A. J., & Hort, J. (2016). Comparing the relative sensitivity of ortho- and retronasal perception of a strawberry flavour model using omission testing. *Flavour and Fragrance Journal*. http://doi.org/10.1002/ffj.3326

Dragulescu, A. A. (2014). xlsx: Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files. Retrieved from http://code.google.com/p/rexcel/, http://groups.google.com/group/R-package-xlsx

Dunkel, A., Steinhaus, M., Kotthoff, M., Nowak, B., Krautwurst, D., Schieberle, P., & Hofmann, T. (2014). Nature's chemical signatures in human olfaction: a foodborne perspective for future biotechnology. *Angewandte Chemie (International Ed. in English)*, *53*(28), 7124–43. http://doi.org/10.1002/anie.201309508

Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *http://dx.doi.org/10.1080/17470218.2012.658821*.

Ennis, D. M. (1993). The Power Of Sensory Discrimination Methods. *Journal of Sensory Studies*, *8*(4), 353–370. http://doi.org/10.1111/j.1745-459X.1993.tb00225.x

Escudero, A., Gogorza, B., Melús, M. A., Ortín, N., Cacho, J., & Ferreira, V. (2004). Characterization of the aroma of a wine from maccabeo. Key role played by compounds with low odor activity values. *Journal of Agricultural and Food Chemistry*, *52*(11), 3516–24. http://doi.org/10.1021/jf035341l

Ferreira, V., Ortín, N., Escudero, A., López, R., & Cacho, J. (2002). Chemical Characterization of the Aroma of Grenache Rosé Wines: Aroma Extract Dilution Analysis, Quantitative Determination, and Sensory Reconstitution Studies. *Journal of Agricultural and Food Chemistry*, *50*(14), 4048–4054. http://doi.org/10.1021/jf0115645

Fox, J., & Weisberg, S. (2011). An R Companion to Applied Regression. Thousands Oaks, CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Gao, W., Fan, W., & Xu, Y. (2014). Characterization of the key odorants in light aroma type chinese liquor by gas chromatography-olfactometry, quantitative measurements, aroma recombination, and omission studies. *Journal of Agricultural and Food Chemistry*, *62*(25), 5796–804. http://doi.org/10.1021/jf501214c

Gottfried, J. A. (2010). Central mechanisms of odour object perception. *Nature Reviews Neuroscience*, *11*(9), 628–641. http://doi.org/10.1038/nrn2883

Grosch, W. (1993). Detection of potent odorants in foods by aroma extract dilution analysis. *Trends in Food Science & Technology*, *4*(3), 68–73. http://doi.org/10.1016/0924-2244(93)90187-F

Grosch, W. (2001). Evaluation of the Key Odorants of Foods by Dilution Experiments, Aroma Models and Omission. *Chemical Senses*, *26*(5), 533–545. http://doi.org/10.1093/chemse/26.5.533

Guth, H., & Grosch, W. (1994). Identification of the Character Impact Odorants of Stewed Beef Juice by Instrumental Analyses and Sensory Studies. *Journal of Agricultural and Food Chemistry*, *42*(12), 2862–2866. http://doi.org/10.1021/jf00048a039

Heeger, D. (1998). Signal Detection Theory. Retrieved August 5, 2016, from http://www-psych.stanford.edu/~lera/psych115s/notes/signal/

Hojsgaard, S., Halekoh, U., & Yan, J. (2006). The R Package geepack for Generalized Estimating Equations journal of Statistical Software.

Hongsoongnern, P. (2003). Learning to Discriminate a Five-Odorant Mixture From Mixtures Containing Four of the Five Odorants.

Howard, J. D., Plailly, J., Grueschow, M., Haynes, J.-D., & Gottfried, J. A. (2009). Odor quality coding and categorization in human posterior piriform cortex. *Nature Neuroscience*, *12*(7), 932–8. http://doi.org/10.1038/nn.2324

Jehl, C., Royet, J. P., & Holley, A. (1995). Odor discrimination and recognition memory as a. *Perception & Psychophysics*, *57*(7), 1002–1011. http://doi.org/10.3758/BF03205459

Jinks, A., & Laing, D. . (2001). The analysis of odor mixtures by humans: evidence for a configurational process. *Physiology & Behavior*, *72*(1-2), 51–63. http://doi.org/10.1016/S0031-9384(00)00407-8

Jinks, A., & Laing, D. G. (1999). A limit in the processing of components in odour mixtures. *Perception*, *28*(3), 395–404. http://doi.org/10.1068/p2898

Kiatbenjakul, P., Intarapichet, K.-O., & Cadwallader, K. R. (2015). Characterization of potent odorants in male giant water bug (Lethocerus indicus Lep. and Serv.), an important edible insect of Southeast Asia. *Food Chemistry*, *168*, 639–647. http://doi.org/10.1016/j.foodchem.2014.07.108

Laing, D. G., & Francis, G. W. (1989). The capacity of humans to identify odors in mixtures. *Physiology & Behavior*, *46*(5), 809–814. http://doi.org/10.1016/0031-9384(89)90041-3

Laing, D. G., & Glemarec, A. (1992). Selective attention and the perceptual analysis of odor mixtures. *Physiology & Behavior*, *52*(6), 1047–1053. http://doi.org/10.1016/0031-9384(92)90458-E

Laing, D. G., & Willcox, M. E. (1983). Perception of components in binary odour mixtures. *Chemical Senses*, *7*(3-4), 249–264. http://doi.org/10.1093/chemse/7.3-4.249

Laing, D., & Jinks, A. (2001). Psychophysical analysis of complex odor mixtures. *CHIMIA International Journal for …*, *55*(5), 413–420. Retrieved from http://www.ingentaconnect.com/content/scs/chimia/2001/00000055/00000005/art00007

Laing, D., Panhuber, H., Willcox, M., & Pittman, E. (1984). Quality and intensity of binary odor mixtures. *Physiology & Behavior*, *33*(2), 309–319. http://doi.org/10.1016/0031-9384(84)90118-5

Lawless, H. T., & Heymann, H. (2010). *Sensory Evaluation of Food: Principles and Practices* (second). Springer, NY: Springer Science and Business Media.

Le Berre, E., Béno, N., Ishii, A., Chabanet, C., Etiévant, P., & Thomas-Danguin, T. (2008). Just noticeable differences in component concentrations modify the odor quality of a blending mixture. *Chemical Senses*, *33*(4), 389–95. http://doi.org/10.1093/chemse/bjn006

Le Berre, E., Thomas-Danguin, T., Béno, N., Coureaud, G., Etiévant, P., & Prescott, J. (2008). Perceptual processing strategy and exposure influence the perception of odor mixtures. *Chemical Senses*, *33*(2), 193–9. http://doi.org/10.1093/chemse/bjm080

Livermore, A., & Laing, D. G. (1996). Influence of training and experience on the perception of multicomponent odor mixtures. *Journal of Experimental Psychology*, *22*(2), 267–277. Retrieved from http://psycnet.apa.orgjournals/xhp/22/2/267

Livermore, A., & Laing, D. G. (1998a). The influence of chemical complexity on the perception of multicomponent odor mixtures. *Perception & Psychophysics*, *60*(4), 650–661. http://doi.org/10.3758/BF03206052

Livermore, A., & Laing, D. G. (1998b). The influence of odor type on the discrimination and identification of odorants in multicomponent odor mixtures. *Physiology & Behavior*, *65*(2), 311–320. http://doi.org/10.1016/S0031-9384(98)00168-1

Lovitz, A. M., Sloan, A. M., Rennaker, R. L., & Wilson, D. A. (2012). Complex mixture discrimination and the role of contaminants. *Chemical Senses*, *37*(6), 533–40. http://doi.org/10.1093/chemse/bjs006

Malnic, B., Hirono, J., Sato, T., Buck, L. B., Adrian, E. ., Arctander, S., … Firestein, S. (1999). Combinatorial receptor codes for odors. *Cell*, *96*(5), 713–23. http://doi.org/10.1016/s0092-8674(00)80581-4

Mayer, F., Czerny, M., & Grosch, W. (2000). Sensory study of the character impact aroma compounds of a coffee beverage. *European Food Research and Technology*, *211*(4), 272–276. http://doi.org/10.1007/s002170000169

Mayr, C. M., Geue, J. P., Holt, H. E., Pearson, W. P., Jeffery, D. W., & Francis, I. L. (2014). Characterization of the key aroma compounds in Shiraz wine by quantitation, aroma reconstitution, and omission studies. *Journal of Agricultural and Food Chemistry*, *62*(20), 4528–36. http://doi.org/10.1021/jf405731v

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (2016). *Sensory Evaluation Techniques* (5th ed.). Boca Raton, FL: Taylor & Francis Group.

Olsson, M. J. (1994). An interaction model for odor quality and intensity. *Perception & Psychophysics*, *55*(4), 363–372. http://doi.org/10.3758/BF03205294

Olsson, M. J. (1998). An Integrated Model of Intensity and Quality of Odor Mixtures. *Annals of the New York Academy of Sciences*, *855*(1 Olfaction and), 837–840. http://doi.org/10.1111/j.1749-6632.1998.tb10672.x

Pang, X., Chen, D., Hu, X., Zhang, Y., & Wu, J. (2012). Verification of aroma profiles of Jiashi muskmelon juice characterized by odor activity value and gas chromatography-olfactometry/detection frequency analysis: aroma reconstitution experiments and omission tests. *Journal of Agricultural and Food Chemistry*, *60*(42), 10426–32. http://doi.org/10.1021/jf302373g

Pavez, C., Steinhaus, M., Casaubon, G., Schieberle, P., & Agosin, E. (2015). Identification, quantitation and sensory evaluation of methyl 2- and methyl 3-methylbutanoate in varietal red wines. *Australian Journal of Grape and Wine Research*, *21*(2), n/a–n/a. http://doi.org/10.1111/ajgw.12131

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

Rabin, M. D. (1988). Experience facilitates olfactory quality discrimination. *Perception & Psychophysics*, *44*(6), 532–540. http://doi.org/10.3758/BF03207487

Reiners, J., & Grosch, W. (1998). Odorants of Virgin Olive Oils with Different Flavor Profiles. *Journal of Agricultural and Food Chemistry*, *46*(7), 2754–2763. http://doi.org/10.1021/jf970940b

Schieberle, P., Gassenmeier, K., Guth, H., Sen, A., & Grosch, W. (1993). Character Impact Odour Compounds of Different Kinds of Butter. *LWT - Food Science and Technology*, *26*(4), 347–356. http://doi.org/10.1006/fstl.1993.1070

Schieberle, P., & Hofmann, T. (1997). Evaluation of the Character Impact Odorants in Fresh Strawberry Juice by Quantitative Measurements and Sensory Studies on Model Mixtures. *Journal of Agricultural and Food Chemistry*, *45*(1), 227–232. http://doi.org/10.1021/jf960366o

Sinding, C., Thomas-Danguin, T., Chambault, A., Béno, N., Dosne, T., Chabanet, C., … Coureaud, G. (2013a). Rabbit neonates and human adults perceive a blending 6-component odor mixture in a comparable manner. *PloS One*, *8*(1), e53534. http://doi.org/10.1371/journal.pone.0053534

Sinding, C., Thomas-Danguin, T., Chambault, A., Béno, N., Dosne, T., Chabanet, C., … Coureaud, G. (2013b). Rabbit neonates and human adults perceive a blending 6-component odor mixture in a comparable manner. *PloS One*, *8*(1), e53534. http://doi.org/10.1371/journal.pone.0053534

Smith, J. D., Zakrzewski, A. C., Herberger, E. R., Boomer, J., Roeder, J. L., Ashby, F. G., & Church, B. A. (2015). The time course of explicit and implicit categorization. *Attention, Perception, & Psychophysics*, *77*(7), 2476–2490. http://doi.org/10.3758/s13414-015-0933-2

Soto, F. A., & Ashby, F. G. (2015). Categorization training increases the perceptual separability of novel dimensions. *Cognition*, *139*, 105–129. http://doi.org/10.1016/j.cognition.2015.02.006

Steinhaus, M., Sinuco, D., Polster, J., Osorio, C., & Schieberle, P. (2009). Characterization of the key aroma compounds in pink guava (Psidium guajava L.) by means of aroma re-engineering experiments and omission tests. *Journal of Agricultural and Food Chemistry*, *57*(7), 2882–8. http://doi.org/10.1021/jf803728n

Stevenson, R. J. (2001). Perceptual learning with odors: Implications for psychological accounts of odor quality perception. *Psychonomic Bulletin & Review*, *8*(4), 708–712. http://doi.org/10.3758/BF03196207

Stevenson, R. J., & Boakes, R. A. (2003). A mnemonic theory of odor perception. *Psychological Review*, *110*(2), 340–364. Retrieved from http://psycnet.apa.orgjournals/rev/110/2/340

Stevenson, R. J., & Mahmut, M. K. (2013). Familiarity influences odor memory stability. *Psychonomic Bulletin & Review*, *20*(4), 754–9. http://doi.org/10.3758/s13423-013-0380-9

Stevenson, R. J., & Wilson, D. A. (2007a). Odour perception: An object-recognition approach. *Perception*, *36*(12), 1821–1833. http://doi.org/10.1068/p5563

Stevenson, R. J., & Wilson, D. A. (2007b). Odour Perception: An Object-Recognition Approach. *Perception*, *36*(12), 1821–1833. http://doi.org/10.1068/p5563

Sweeten, J. M., McFarland, A. R., Sorel, J. E., Gauntt, R. O., & Redded, D. L. (1984). A Butanol Olfactometer for Field Measurement of Ambient Odors. *Journal of the Air Pollution Control Association*, *34*(12), 1208–1213. http://doi.org/10.1080/00022470.1984.10465872

Tokitomo, Y., Steinhaus, M., Buttner, A., & Schieberle, P. (2014). Odor-Active Constituents in Fresh Pineapple (Ananas comosus [L.] Merr.) by Quantitative and Sensory Evaluation. *Bioscience, Biotechnology and Biochemistry*, *69*(7), 1323–1330. http://doi.org/10.1271/bbb.69.1323

Wagner, R. K., & Grosch, W. (1998). Key odorants of french fries. *Journal of the American Oil Chemists' Society*, *75*(10), 1385–1392. http://doi.org/10.1007/s11746-998-0187-4

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research, A*(2), 149–168. Retrieved from http://statpages.info/latinsq.html

Wilson, D. A. (2003). Rapid, experience-induced enhancement in odorant discrimination by anterior piriform cortex neurons. *Journal of Neurophysiology*, *90*(1), 65–72. http://doi.org/10.1152/jn.00133.2003

Wilson, D. A., & Sullivan, R. M. (2011). Cortical processing of odor objects. *Neuron*, *72*(4), 506–19. http://doi.org/10.1016/j.neuron.2011.10.027

Zou, Z., & Buck, L. B. (2006). Combinatorial Effects of Odorant Mixes in Olfactory Cortex. *Science*, *311*(5766), 1477–1481. Retrieved from http://www.jstor.org.ezp1.lib.umn.edu/stable/3845788

Appendices

Appendix A: The recruitment email

Subject: Recruiting Panelists for Sensory Evaluation of Odor Mixtures

Hello,

My name is Kirsten Weiss and I am a graduate student in the Food Science and Nutrition Department at the University of Minnesota. I am looking for panelists who are willing to participate in my research study on Sensory Evaluation of Odor Mixtures.

This study requires panelists who are willing to make a commitment and are able to attend between 2 and 25 sessions during the next few months. Each session will involve smelling various odors and answering questions regarding topics such as intensity, familiarity, and identity. Panelists will be compensated $5 per session. Sessions will be held in McNeal Hall on the St. Paul campus of the University of Minnesota and will be approximately 30 minutes long.

If you are interested in being a panelist and can make the commitment, please contact me at weis0419@umn.edu.

Thank you,

Kirsten Weiss

Appendix B: Qualtrics survey used for recruiting panelists

Sensory Evaluation of Odor Mixtures

I am recruiting MEN and WOMEN for a study on the sensory evaluation of odor mixtures. This study will be conducted in Room 132 of McNeal Hall on the St. Paul Campus. To participate in the study I would ask that you be willing to attend between 2 and 25 sessions during the next few months. Each session would be approximately 30 minutes long and would involve smelling various odors and answering questions. All sessions are mandatory. However, I can be flexible with scheduling sessions. Each participant will be compensated $5 per session. If you are interested in taking part in this study, please answer ALL of the questions in this survey. I will contact you in the next few days to schedule you for the study.

Please enter your email address:

Please indicate your availability for sessions during the week of August 24th through August 28th.

- ❍ I am available to attend sessions during this week
- ❍ I am not available this week, but I am interested in participating in the study.

Please indicate the times on Monday, August 24th that you would be able to attend a session. (Please mark all available times!)

- ❑ 8:00AM
- ❑ 8:30AM
- ❑ 9:00AM
- ❑ 9:30AM
- ❑ 10:00AM
- ❑ 10:30AM
- ❑ 11:00AM
- ❑ 11:30AM
- ❑ 12:00PM
- ❑ 12:30PM
- ❑ 1:00PM
- ❑ 1:30PM
- ❑ 2:00PM
- ❑ 2:30PM
- ❑ 3:00PM
- ❑ I am not available this day

Please indicate the times on Tuesday, August 25th that you would be able to attend a session. (Please mark all available times!)

- ❑ 8:00AM
- ❑ 8:30AM
- ❑ 9:00AM
- ❑ 9:30AM
- ❑ 10:00AM
- ❑ 10:30AM
- ❑ 11:00AM
- ❑ 11:30AM
- ❑ 12:00PM
- ❑ 12:30PM
- ❑ 1:00PM
- ❑ 1:30PM
- ❑ 2:00PM
- ❑ 2:30PM
- ❑ 3:00PM
- ❑ I am not available this day

Please indicate the times on Thursday, August 27th that you would be able to attend a session. (Please mark all available times!)

- ❑ 8:00AM
- ❑ 8:30AM
- ❑ 9:00AM
- ❑ 9:30AM
- ❑ 10:00AM
- ❑ 10:30AM
- ❑ 11:00AM
- ❑ 11:30AM
- ❑ 12:00PM
- ❑ 12:30PM
- ❑ 1:00PM
- ❑ 1:30PM
- ❑ 2:00PM
- ❑ 2:30PM
- ❑ 3:00PM
- ❑ I am not available this day

Please indicate the times on Friday, August 28th that you would be able to attend a session. (Please mark all available times!)

- ❑ 8:00AM
- ❑ 8:30AM
- ❑ 9:00AM
- ❑ 9:30AM
- ❑ 10:00AM
- ❑ 10:30AM
- ❑ 11:00AM
- ❑ 11:30AM
- ❑ 12:00PM
- ❑ 12:30PM
- ❑ 1:00PM
- ❑ 1:30PM
- ❑ 2:00PM
- ❑ 2:30PM
- ❑ 3:00PM
- ❑ I am not available this day

Thank you for your interest in the Sensory Evaluation of Odor Mixtures study! I will contact you in the next few days to schedule you for the study. You may choose not to participate even if you have completed this survey. If you have any questions about the study please respond to this e-mail or contact me at weis0419@umn.edu.

Appendix C: Consent Form

CONSENT FORM

Sensory Evaluation of Odor Mixtures

You are invited to be in a research study on sensory evaluation of odor mixtures. You were selected as a possible participant because of your interest and availability. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Kirsten Weiss from the Department of Food Science and Nutrition.

Procedures:

If you agree to be in this study, we would ask you to participate in somewhere between 2 and 25 sessions during which you will smell various odors and answer questions about them regarding topics such as intensity, familiarity, and identity. Sessions will be approximately 30 minutes long.

Compensation:

You will receive $5 per session.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher conducting this study is: Kirsten Weiss. You may ask any questions you have now. If you have questions later, you are encouraged to contact Kirsten at weis0419@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Appendix D: Summary of the results of the intensity matching phase for all panelists for **a)** butyric acid, **b)** methional, **c)** furaneol, and **d)** δ-decalactone. The values for intercept, slope, and R^2 were obtained from the regression line for intensity rating vs concentration. Concentration is the concentration matching the intensity of the standard obtained from the equation for the regression line.

a) Butyric Acid

Panelist	Intercept	Slope	R^2	Concentration Matching Standard (ppm)
1	42.82	0.24	0.58	132.51
2	37.41	0.46	0.52	82.11
3	34.97	0.37	0.48	107.17
4	54.25	0.21	0.29	101.20
5	55.61	0.21	0.19	93.02
6	47.44	0.27	0.73	103.28
7	51.69	0.23	0.46	103.07
8	36.12	0.12	0.10	329.30
9	87.98	0.14	0.15	-94.64
10	44.58	0.28	0.55	109.08
11	14.97	0.43	0.75	140.62
12	32.01	0.33	0.58	128.74
13	26.74	0.48	0.76	100.12
14	21.28	0.50	0.68	107.42
15	22.39	0.52	0.80	101.41
16	16.35	0.37	0.64	159.35
17	26.80	0.60	0.66	80.01
18	33.12	0.23	0.55	181.30
19	38.66	0.61	0.67	59.19
20	49.21	0.45	0.49	57.86
21	11.81	0.34	0.77	188.46
22	36.58	0.49	0.70	78.10
23	30.80	0.58	0.70	76.40
24	28.34	0.28	0.41	166.27
25	33.37	0.28	0.45	147.61
26	18.42	0.40	0.72	142.19
27	55.73	0.23	0.49	84.61
28	21.98	-0.01	0.01	-3918.95
29	71.29	0.34	0.76	11.02
30	41.15	0.38	0.48	89.66
31	65.10	0.45	0.49	21.99
32	20.34	0.37	0.56	146.17
33	43.21	0.15	0.24	207.16
34	22.01	0.44	0.61	120.37
35	23.50	0.28	0.66	186.77

b) Methional

Panelist	Intercept	Slope	R^2	Concentration Matching Standard (ppm)
1	89.53	11.29	0.76	0.276
2	109.07	5.79	0.33	0.003
3	78.60	8.38	0.49	0.65
4	110.76	7.48	0.69	0.01
5	80.97	6.49	0.57	0.074
6	114.19	6.61	0.76	0.003
7	104.88	9.54	0.77	0.044
8	104.82	9.61	0.69	0.045
9	127.51	9.62	0.77	0.004
10	113.18	12.64	0.92	0.049
11	93.32	12.92	0.62	0.242
12	92.59	10.65	0.94	0.192
13	108.03	14.44	0.72	0.101
14	102.19	15.85	0.92	0.180
15	113.24	14.60	0.93	0.073
16	86.11	11.99	0.85	0.396
17	127.56	17.47	0.85	0.049
18	91.88	3.46	0.43	0.008
19	107.83	12.62	0.83	0.074
20	107.54	15.20	0.88	0.117
21	110.44	15.46	0.92	0.101
22	85.66	5.30	0.54	0.199
23	108.31	13.58	0.58	0.086
24	36.48	6.01	0.68	607.068
25	74.48	10.45	0.57	1.051
26	82.00	14.69	0.81	0.621
27	92.60	12.22	0.87	0.237
28	77.72	9.97	0.84	0.761
29	118.73	12.91	0.79	0.034
30	97.90	9.44	0.70	0.088
31	124.16	12.75	0.78	0.021
32	40.51	3.62	0.13	13758.538
33	34.43	-5.23	0.25	0.000
34	92.20	4.81	0.69	0.028
35	139.12	6.32	0.59	0.000

c) Furaneol

Panelist	Intercept	Slope	R^2	Concentration Matching Standard (ppm)
1	43.14	0.31	0.38	102.73
2	13.99	0.36	0.42	169.08
3	39.00	0.21	0.31	174.71
4	30.83	0.20	0.23	224.59
5	64.38	0.05	0.21	215.57
6	27.83	0.20	0.41	238.91
7	35.61	0.24	0.45	167.02
8	25.58	0.26	0.60	191.74
9	50.12	0.22	0.45	114.15
10	49.30	0.10	0.36	253.95
11	36.87	0.28	0.74	137.32
12	57.30	0.05	0.24	324.85
13	9.30	0.20	0.77	327.99
14	18.70	0.12	0.44	472.23
15	59.32	0.09	0.22	167.16
16	25.18	0.16	0.55	318.23
17	37.35	0.21	0.49	177.56
18	32.29	0.08	0.33	513.84
19	16.37	0.59	0.66	99.63
20	30.46	0.45	0.77	99.80
21	34.05	0.32	0.63	129.49
22	48.39	0.20	0.25	136.00
23	67.46	0.22	0.49	34.01
24	8.82	0.04	0.09	1780.11
25	0.16	0.00	0.07	80217.72
26	-3.91	0.20	0.42	393.33
27	60.78	0.06	0.05	227.21
28	8.96	0.02	0.04	2873.57
29	50.28	0.01	0.00	4907.95
30	34.23	0.21	0.47	195.52
31	21.14	0.04	0.03	1256.01
32	-4.20	0.15	0.70	533.87
33	38.80	0.02	0.03	1711.76
34	50.57	-0.01	0.01	-1796.67
35	38.79	0.12	0.38	305.08

d) δ-decalactone

Panelist	Intercept	Slope	R^2	Concentration Matching Standard (ppm)
1	24.41	1.03	0.88	49.09
2	46.30	0.49	0.15	58.71
3	57.82	0.30	0.30	57.98
4	57.06	0.26	0.46	68.40
5	62.39	0.42	0.42	29.70
6	28.19	0.37	0.62	125.93
7	54.35	0.28	0.40	73.73
8	31.44	0.47	0.61	93.48
9	60.26	0.39	0.64	38.11
10	50.72	0.29	0.61	83.23
11	67.41	0.27	0.19	28.44
12	50.85	0.23	0.46	103.10
13	61.06	0.31	0.53	44.42
14	38.45	0.32	0.52	113.93
15	54.07	0.18	0.16	118.86
16	20.09	0.33	0.51	168.44
17	33.74	0.37	0.66	111.72
18	59.91	0.07	0.35	202.25
19	28.45	1.26	0.75	37.01
20	28.87	0.25	0.34	188.05
21	61.67	0.36	0.53	36.67
22	57.47	0.25	0.43	70.64
23	119.24	0.12	0.18	-358.26
24	14.43	0.44	0.51	138.97
25	69.75	0.07	0.05	79.10
26	37.48	0.74	0.62	50.96
27	72.36	0.33	0.12	8.00
28	12.39	0.25	0.32	248.76
29	54.53	0.36	0.48	57.01
30	10.30	0.23	0.35	286.02
31	47.91	0.80	0.75	34.00
32	43.49	0.21	0.28	148.59
33	43.40	0.27	0.22	117.94
34	36.49	0.13	0.13	303.69
35	78.64	0.17	0.21	-21.27

Appendix E: Sample codes used in the intensity matching phase

	Session 1			Session 2		
	Sample (ppm)	**Sample**	**Code**	**Sample (ppm)**	**Sample**	**Code**
Butyric Acid	0.1	S1	826	0.01	S1	321
Session 1: Test 472	0.1	S2	569	0.01	S2	263
Session 2: Test 128	50	S3	152	1	S3	416
	50	S4	653	1	S4	625
Yellow Tray	100	S5	593	100	S5	796
ARoom1	100	S6	912	100	S6	567
Kirsten1BA	150	S7	453	125	S7	382
	150	S8	369	125	S8	146
	200	S9	652	250	S9	485
	200	S10	183	250	S10	967
Methional	0.001	S1	792	0.001	S1	426
Session 1: Test 683	0.001	S2	189	0.001	S2	217
Session 2: Test 593	0.01	S3	382	0.01	S3	152
	0.01	S4	135	0.01	S4	615
	0.1	S5	476	0.1	S5	864
Green Tray	0.1	S6	381	0.1	S6	984
ARoom2	1	S7	916	1	S7	324
Kirsten1M	1	S8	796	1	S8	753
	10	S9	361	10	S9	617
	10	S10	684	10	S10	746
δ-Decalactone	0.1	S1	748	0.01	S1	597
Session 1: Test 325	0.1	S2	529	0.01	S2	674
Session 2: Test 921	25	S3	276	10	S3	892
	25	S4	943	10	S4	238
White Tray	50	S5	738	75	S5	642
ARoom3	50	S6	519	75	S6	731
Kirsten1DD	75	S7	359	150	S7	349
	75	S8	136	150	S8	683
	100	S9	974	200	S9	129
	100	S10	489	200	S10	694

Furaneol **Session 1: Test 108** **Session 2: Test 748** **Red Tray** **ARoom4** **Kirsten1F**	0.1	S1	413	1	S1	678
	0.1	S2	986	1	S2	582
	50	S3	715	75	S3	791
	50	S4	958	75	S4	192
	100	S5	839	150	S5	826
	100	S6	925	150	S6	953
	150	S7	267	225/275	S7	768
	150	S8	318	225/275	S8	945
	200	S9	146	300/500	S9	872
	200	S10	273	300/500	S10	278

Appendix F: SIMS Sensory Software instructions for intensity matching.

Sensory Evaluation of Odor Mixtures
Welcome!

Today you will be rating the intensity of different odor samples in comparison to a standard. Please focus only on the intensity of the samples, not the odor characteristics.

You will be rating 4 sets of samples. There will be a 30 second wait between samples and a 2 minute wait between sets.

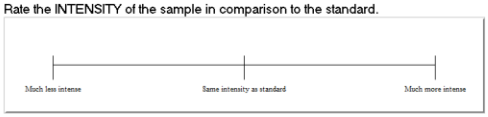
Click the icon on the top of the screen to proceed.

Gently swirl the standard or sample before smelling.
To smell, please remove the lid, leave the bottle on the counter, and lower your head to the sample. Remember to replace the lid when finished.
Smell the standard first and remember its intensity. Smell the sample and rate its intensity in comparison to that of the standard by placing a mark on the scale below.

Smell the standard before evaluating EVERY sample.

Please make sure the sample you are evaluating corresponds to the number on the screen.

Rate the INTENSITY of the sample in comparison to the standard.



Click the icon on the top of the screen to proceed.

Please wait 30 seconds before rating the next sample.

Remember to smell the standard first!

30

You have completed this set. Slide your tray through the window to receive the next set of samples. Please wait 2 minutes before beginning the next set. If this was your 4th set, you have completed today's session. Thank you for participating!

120

Appendix G: SAS Code for Phase 1: Intensity Matching

```
libname xxx 'G:\FSCN\Vickers_Lab\Kirsten Weiss\Thesis\Analysis\SAS data sets';
data xxx.butyric;
set butyric;
run;
proc print data=xxx.butyric;
run;
/*run data separately for each individual judge*/
proc sort data=xxx.butyric;
by judge;
run;
/*proc reg = regression analysis, outest = outputs data set that contains specified
parameters, regstat = file name
given to results, TABLEOUT = adds estimates for p-vlaue t-stat etc, EDF adds R-squared
value to output*/
proc reg data=xxx.butyric outest=regstat TABLEOUT EDF;
by judge;
/*create the regression model y=x */
model intensity = conc;
run;
/*try log values */
/*create new data set*/
data xxx.butyriclog;
/*from this original set*/
set xxx.butyric;
/*with a new variable = logconc */
logconc = log(conc);
run;
/*once again run for each individual judge*/
proc sort data=xxx.butyriclog;
by judge;
run;
/*run regression on model with logconc */
proc reg data=xxx.butyriclog outest=regstatL TABLEOUT EDF;
by judge;
model intensity = logconc;
run;
libname xxx 'G:\FSCN\Vickers_Lab\Kirsten Weiss\Thesis\Analysis\SAS data sets';
data xxx.deltaD;
set deltaD;
run;
proc print data=xxx.deltaD;
run;
```

```
/*run data separately for each individual judge*/
proc sort data=xxx.deltaD;
by judge;
run;
/*proc reg = regression analysis, outest = outputs data set that contains specified
parameters, regstat = file name given to results, TABLEOUT = adds estimates for p-
vlaue t-stat etc, EDF adds R-squared value to output*/
proc reg data=xxx.deltaD outest=regstat TABLEOUT EDF;
by judge;
/*create the regression model y=x */
model intensity = conc;
run;
/*try log values */
/*create new data set*/
data xxx.deltaDlog;
/*from this original set*/
set xxx.deltaD;
/*with a new variable = logconc */
logconc = log(conc);
run;
/*once again run for each individual judge*/
proc sort data=xxx.deltaDlog;
by judge;
run;
/*run regression on model with logconc */
proc reg data=xxx.deltaDlog outest=regstatL TABLEOUT EDF;
by judge;
model intensity = logconc;
run;
libname xxx 'G:\FSCN\Vickers_Lab\Kirsten Weiss\Thesis\Analysis\SAS data sets';
data xxx.methional;
set methional;
run;
proc print data=xxx.methional;
run;
/*run data separately for each individual judge*/
proc sort data=xxx.methional;
by judge;
run;
/*proc reg = regression analysis, outest = outputs data set that contains specified
parameters, regstat = file name given to
results, TABLEOUT = adds estimates for p-vlaue t-stat etc, EDF adds R-squared value to
output*/
```

```
proc reg data=xxx.methional outest=regstat TABLEOUT EDF;
by judge;
/*create the regression model y=x */
model intensity = conc;
run;
/*try log values*/
/*create new data set*/
data xxx.methionallog;
/*from this original set*/
set xxx.methional;
/*with a new variable = logconc */
logconc = log(conc);
run;
/*once again run for each individual judge*/
proc sort data=xxx.methionallog;
by judge;
run;
/*run regression on model with logconc */
proc reg data=xxx.methionallog outest=regstatL TABLEOUT EDF;
by judge;
model intensity = logconc;
run;
libname xxx 'G:\FSCN\Vickers_Lab\Kirsten Weiss\Thesis\Analysis\SAS data sets';
data xxx.furaneol;
set furaneol;
run;
proc print data=xxx.furaneol;
run;
/*run data separately for each individual judge*/
proc sort data=xxx.furaneol;
by judge;
run;
/*proc reg = regression analysis, outest = outputs data set that contains specified
parameters, regstat = file name given to results, TABLEOUT = adds estimates for p-
vlaue t-stat etc, EDF adds R-squared value to output*/
proc reg data=xxx.furaneol outest=regstat TABLEOUT EDF;
by judge;
/*create the regression model y=x */
model intensity = conc;
run;
/*try log values */
/*create new data set*/
data xxx.furaneollog;
/*from this original set*/
```

```
set xxx.furaneol;
/*with a new variable = logconc */
logconc = log(conc);
run;
/*once again run for each individual judge*/
proc sort data=xxx.furaneollog;
by judge;
run;
/*run regression on model with logconc */
proc reg data=xxx.furaneollog outest=regstatL TABLEOUT EDF;
by judge;
model intensity = logconc;
run;
```

Appendix H: Omission testing sample codes for a) sessions 1 – 10 and b) sessions 11 - 20.

a)	Session									
Sample	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	724	856	684	715	267	832	387	892	981	975
2	863	275	197	586	143	978	861	248	453	251
3	261	612	538	194	286	598	623	538	789	396
4	824	289	136	826	512	167	714	621	267	254
5	936	918	615	756	815	719	163	962	478	359
6	421	769	514	192	945	682	519	134	125	847
7	975	296	362	825	175	536	716	937	316	921
8	463	846	975	158	923	874	495	843	437	326
9	837	324	635	952	863	713	123	245	821	715
10	291	975	265	524	452	546	785	495	149	659
11	954	761	164	976	943	497	314	958	752	298
12	354	453	872	293	798	537	621	738	859	763
13	849	358	518	396	381	137	813	425	592	864
14	563	713	276	893	784	534	954	793	791	284
15	925	983	375	736	364	296	795	857	516	978
16	432	875	285	356	957	572	162	594	815	217
17	719	298	638	839	345	437	239	219	539	638
18	146	834	986	342	615	749	743	137	957	769
19	912	386	649	589	871	295	185	746	495	596
20	394	874	497	796	532	482	359	173	765	821

b)	Session									
Sample	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
1	918	243	831	872	736	234	831	521	267	236
2	613	356	351	978	249	853	128	486	916	432
3	257	815	461	265	792	928	796	724	543	249
4	638	162	126	195	514	684	231	258	628	629
5	752	738	452	941	413	528	861	489	954	461
6	537	179	234	724	617	376	732	234	136	916
7	768	693	543	892	259	896	947	371	978	129
8	843	143	924	912	437	624	725	719	739	674
9	759	352	851	786	187	478	481	928	348	197
10	127	132	593	523	396	389	317	137	826	731
11	581	256	769	174	964	594	721	287	148	416
12	984	597	684	358	258	314	817	685	635	145
13	231	784	784	138	832	592	372	528	365	571
14	524	315	524	713	342	752	784	741	269	871
15	927	514	156	517	529	358	518	241	736	624
16	468	745	873	842	748	925	843	735	972	298
17	643	517	492	231	427	156	397	917	596	589
18	589	479	916	743	268	248	826	742	186	854
19	837	916	287	528	438	647	769	543	932	367
20	689	187	594	821	762	173	259	125	192	618

Appendix I: SIMS Sensory Software instructions for omission testing

Appendix J: R code for omission testing analysis

```
library(geepack)
library(xlsx)
library(car)
library(ggplot2)

##### data processing: judge, rsp01, session: numeric; num.removed: factor
getwd()  # the following outputs are saved in the default working director, whose path
will be returned by this command
d1 <- read.csv("discriminating2.csv")
d2 <- read.csv("samples.csv")
d <- merge(d1, d2)
rm(list=c("d1","d2"))

d$response <- d$response - 1L  # 1: not A; 0: A
d$num.removed <- as.factor(d$num.removed)  # to factorize num.removed
d$which.removed <- as.factor(d$which.removed)  # to factorize which.removed
d_ms <- d[,c("response","judge","session","order","num.removed","which.removed")]  #
to extract the columns we need

# to have subject-by-session combination
d_ms$judgesession <- as.factor(paste(d_ms$judge, d_ms$session))
d_ms <- d_ms[with(d_ms, order(judge, session, order)),]


# written functions to output results in a desired form
addci <- function(m) {
  cc <- summary(m)$coefficients
  k <- qnorm(1-0.05/2)
  within(cc, {
    upper <- Estimate + k*Std.err
    lower <- Estimate - k*Std.err
  })
}

extractone <- function(m, about) {
  cc <- addci(m)
  cc <- cc[grepl(about, rownames(cc)),]  # grepl() for each element in character list
  tmp <- rownames(cc)
  tmp <- sub(about,"",tmp, fixed=TRUE)
  out <- data.frame(about=tmp, cc, check.names=FALSE)
  rownames(out) <- NULL
  out
```

```
}
extractit <- function(m, var, about) {
  cc <- addci(m)
  cc <- cc[grepl(var, rownames(cc)),]
  tmp <- rownames(cc)
  tmp <- sub(about,"",tmp, fixed=TRUE)
  tmp <- sub(paste0(":",var), "", tmp, fixed=TRUE)
  out <- data.frame(variable=var, about=tmp, cc, check.names=FALSE)
  rownames(out) <- NULL
  out
}


##########1. Primary goal: N-1 vs Full blend, N-2 vs Full blend;
###    All the sessions are included

### 1.1 dprimes for sessions
m_gee_mn11 <- geeglm(response ~ factor(session) + factor(session) : num.removed,
family=binomial(link = "probit"),
                data = d_ms, id=judgesession, corstr = "exchangable")

out11 <- rbind(extractit(m_gee_mn11, "num.removed1", "factor(session)"),
            extractit(m_gee_mn11, "num.removed2", "factor(session)"))
out11$about <- as.integer(as.character(out11$about))

plot11 <- ggplot(out11) + aes(about, Estimate, ymin=lower, ymax=upper, color=variable)
+
  geom_pointrange(position=position_dodge(width=0.3)) + xlab("Session")

# to output the .csv
write.csv(out11,file="DPrime_session.csv")


### 1.2 dprimes for num.removed
m_gee_mn12 <- geeglm(response ~ num.removed, family=binomial(link = "probit"),
                data = d_ms, id=judgesession, corstr = "exchangable")

out12 <- extractone(m_gee_mn12, "num.removed")
plot12 <- ggplot(out12) + aes(about, Estimate, ymin=lower, ymax=upper) +
geom_pointrange() + xlab("num removed")

# to output the .csv
write.csv(out12,file="DPrime_NumRemoved.csv")
```

```
###### 2.Secondary goal: compounds; judge
d_10 <- d_ms[d$session>=11,] # take the last 10 sessions

### 2.1 judge: across last 10 sessions, across N-1
m_gee_mn21 <- geeglm(response ~ as.factor(judge) + as.factor(judge):num.removed,
family=binomial(link = "probit"),
                data = d_10, id=judgesession, corstr = "exchangable")

out21 <- rbind(extractit(m_gee_mn21, "num.removed1", "as.factor(judge)"),
            extractit(m_gee_mn21, "num.removed2", "as.factor(judge)"))
out21$about <- reorder(out21$about, out21$Estimate, mean)

plot21 <- ggplot(out21) + aes(about, Estimate, ymin=lower, ymax=upper, color=variable)
+
  geom_pointrange(position=position_dodge(width=0.3)) + xlab("Judge")

# to output the .csv
write.csv(out21,file="DPrimes_Judge_N1.csv")

### 2.2 compound: across last 10 sessions, across N-1
m_gee_mn22 <- geeglm(response ~ which.removed, family=binomial(link = "probit"),
                data = d_10, id=judgesession, corstr = "exchangable")

out22 <- extractone(m_gee_mn22, "which.removed")
plot22 <- ggplot(out22) + aes(about, Estimate, ymin=lower, ymax=upper) +
geom_pointrange() + ylab("which removed")

# to output the .csv
write.csv(out22,file="DPrimes_Compound_N1.csv")


### 2.3 dprimes for num.removed with the last 10 sessions
m_gee_mn23 <- geeglm(response ~ num.removed, family=binomial(link = "probit"),
                data = d_10, id=judgesession, corstr = "exchangable")

out23 <- extractone(m_gee_mn23, "num.removed")
plot23 <- ggplot(out23) + aes(about, Estimate, ymin=lower, ymax=upper) +
geom_pointrange() + ylab("num removed")

# to output the .csv
write.csv(out23,file="DPrime_NumRemoved_L10.csv")
```