

Design and analysis of choice experiments involving mixtures

Mario Becerra

Supervisor:
Prof. dr. Peter Goos

Dissertation presented in partial
fulfillment of the requirements for the
degree of Doctor of Bioscience
Engineering (PhD)

January 2024

Design and analysis of choice experiments involving mixtures

Mario BECERRA

Examination committee:

Prof. dr. ir. Bert Sels, chair

Prof. dr. Peter Goos, supervisor

Prof. dr. Martina Vandebroek

Prof. dr. Michel Meulders

Dr. ir. Bart De Ketelaere

Prof. dr. Roselinde Kessels

(Maastricht University)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Bioscience Engineering (PhD)

© 2024 KU Leuven – Faculty of Bioscience Engineering
Uitgegeven in eigen beheer, Mario Becerra, Kasteelpark Arenberg 30, 3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Acknowledgments

I would like to thank my supervisor, Peter Goos, for being a great mentor during these four years. His encouragement, comradeship, and understanding helped a lot in the course of this PhD journey of mine. Not only his experience and expertise, but also his empathy and compassion, were of great help in many situations. I also appreciate the tips and tricks he gave me related to running and cycling, especially around the Leuven area. Turns out he also has a good deal of experience and expertise in this area. I also always valued his Strava kudos, which worked as an additional encouragement to keep working out.

Furthermore, I would like to thank the members of my doctoral committee Martina Vandebroek, Michel Meulders, Bart De Ketelaere, and Roselinde Kessels. Your questions, your constructive comments, and your suggestions improved the quality and the content of this thesis. Special thanks to Martina for accepting to be part of my supervisory committee since the very beginning and attended several of my talks and seminars. And I would also like to thank Eddie Schrevens for being part of my supervisory committee at the beginning of my PhD. I also thank Prof. Bert Sels for chairing my preliminary defense in a professional manner.

I would also like to like to extend thanks to Frank Mathijs, who created the experimental setup for the fruit flies in Chapter 4, the software to automatically count the number of flies, and performed several experiments. I also thank Ya Wang and Yicheng Mao for their help performing the first experiments and the initial transformation of the colors' wavelengths into mixture form. Finally, thanks to Bart De Ketelaere who was involved in this project since the beginning and helped a lot in its conceptualization and execution. Additionally, his Strava kudos were also always welcome.

Preface

I wrote this preface to tell a bit about how I got here and, above all, about the people who took me along the way. For this is an achievement that is not of me alone, but also of the people who were there beside me. And what's the point of life without friends and family.

It's funny, I never really wanted to do a PhD. This might be a weird way to start a PhD thesis preface, but it's the truth. I wasn't opposed to having one either, but it was never really a dream that I had. I first started thinking of PhDs when I started my bachelor in Applied Mathematics in Mexico. A lot of my professors had PhDs, and it was sort of encouraged to do one eventually. I found the idea of a PhD interesting. The idea of working on a certain topic and digging really deep into it was attractive, but at the same time it seemed a little tedious and frightening. Given that I find many topics interesting, the idea of having to work on only a single topic seemed scary and a bit boring. Nevertheless, I never really considered it because my GPA was not the best someone could have. However, when I started my master and my GPA improved, I started considering it.

Even though I had considered the PhD and it was at the back of my mind, I never did any serious research about where to go or what topic to do. The most serious thing I got to do about it was go to a dinner arranged by my friend (and twice boss) Gerson to meet with Horacio, a recently graduated math PhD from Bath, and to basically interrogate him about PhD life.

Some time later, one night around December of 2018, while I was procrastinating the writing of my master thesis, I saw the description of a PhD vacancy at KU Leuven about design of experiments, something I knew very little of. I knew the university and I remembered that I had visited Leuven a couple of years back and I had really liked it. I checked the requirements for the vacancy and, since I had the profile they were looking for and all I had to do was send my CV with a motivation letter, I applied at that very moment and then forgot almost completely about it.

It was two months later when I received an email from Peter Goos saying he wanted to interview me. In the interview would also be present his then post-doc, José Núñez. I guess they had a good impression of me, because in the follow-up email they said I was their top candidate. However, Peter did ask me to research a bit more about experimental design and to read the first chapters of his book to make sure I was interested in the topic. Since this actually seemed to be going

somewhere, I did it and I also read more about Peter's research and a couple of his papers. I did not understand every little detail of them at first, but I understood enough that I got interested in the topic. It also sparked my interest in doing a PhD for real, which led me to look for other universities and opportunities. However, shortly after my second interview I got an offer to join Peter's research group. I had not even sent any other application to any other PhD program, but it seemed like a good opportunity so I accepted the offer. A few months later, in late September, I was packed up and ready to move to Belgium. My girlfriend Tatiana took me to the airport where we met my cousin Maru because she wanted to say goodbye. And so it began.

I remember José was the first person I met in Leuven, he invited me for coffee before I had even started officially working at KU Leuven. Tatiana joined me for a couple of months. I met my colleagues and now friends Alan, Alex, and Saif. During those months I also go to know Valeria, Jirka, Yannis, Ricardo, David, Enrique, and Abril. My friend Coro came to visit for a few weeks and he, Tatiana, and I traveled around Belgium. I saw my friend Alex in Brussels. I also met up a couple of times with my friend Pancho who lived in Offenburg, Germany at the time. He would come to Belgium or I would go to Offenburg. I even attended his master graduation ceremony and party, where I could see some people I had not seen in a while, like Adam, Carla, and Dani. Tatiana and I also had the chance to travel with him and his (then girlfriend and now) wife, Paola, to the beautiful Neuschwanstein Castle in south Germany.

I got to travel to other places during those first few months. For instance, in January 2020, I met up in Sweden with my friend Diana Castañeda, where we took a flight to the Faroe Islands. Around the time, my cousin Erika and her husband Nico moved to Switzerland, where I visited them (with a stop in Offenburg). We went on a skiing trip on my birthday with Vale and Pablo. In late February, my friend Andrea came from England to visit and spend a weekend in Leuven.

What I want to say is that adjusting to life in Belgium was pretty easy at the beginning and I had a blast. However, in early March, COVID lockdown started in Europe. What many thought was gonna be a lockdown of a few weeks turned into several months. That took a bit more of adjusting, especially since Tatiana had gone to Mexico until she could start her master in Belgium eight months later.

Nevertheless, I had the good fortune of being able to work from home and remain safe. I was also lucky to be able to have video calls with my friends and family. I remember a couple of specific ones, like some calls with my cousin Diego who was also in lockdown in Switzerland, and her sister Paulina in the same situation in Copenhagen. We had calls with our other *primos* Andrea, Roby, Ale, and Daniel. I had video calls with the Contreras family. Video calls with the Coapos: Pancho, Edgar, Melissa, and Truax. I also had regular virtual coffees with people from the office. As well as with my friends Saúl, la Clau, and, of course, Tatiana. Sometimes Tatiana and I would watch the same movie at the same time, and be connected through video call and felt like we were doing a usual movie night.

Fortunately, this first lockdown did not last long, and in July of 2020 I was able to do a nice summer tour to go to Lugano, Switzerland and visit Diego. On the way

I also got to stay in Erika and Nico's brand new house, and on the way back I visited Pancho and Paola in Stuttgart, where we would have to do the obligatory visit to Offenburg. Around that time I met my now friends Sofi and Alejo. Sofi would end up breaking my and Tatiana's hearts by moving to Hasselt to pursue her PhD instead of staying in Leuven. We were happy for her, though.

Those days Tatiana was in process of getting her visa to move to Belgium. That was a long process with its own set of obstacles but in the end she managed to get here in August. Another exciting thing about this happening was that my friend from university, Diana Rosales, was also moving to Leuven to start her master thesis. This event also led me to meet Carlos, Victoria and Christian. Europe at the time had the chance to have a few months of normality, until COVID got us all confined to our homes again by winter that year. One good thing that this brought was that Truax had moved to Vienna to do his master, and he came to Leuven to visit for a few months since all his classes were online anyway.

After this, months seemed to go by very quickly. Being at home most of the time with social distancing did not allow many things to happen. However, I did often practice my French with Florence, while she practiced her Spanish. Truax, Diana, Tatiana, and I would often get together to play dominoes or other games. I would go on socially distanced walks with my friends Daria and Nina. I would go running with Truax, and I would go cycling with José and David. Then I would meet Patrick through our mutual connection, Areta, and he would join us in our cycling trips. Patrick would also then somehow convince me to do with him a 215 Km bike ride in November that would take us through Germany, the Netherlands, Belgium, Luxembourg, and France.

During the summer of 2021, we were fortunate enough in Belgium to get our COVID vaccine, with which things started slowly going back to normal. I got the chance to travel to Switzerland again to visit Erika and Nico, and then to travel to Barcelona with Tatiana, Diana and Truax, where I got the chance to see my old colleague, Citlalli, and my old classmate, Eduardo. I then traveled to Madrid with Truax and Diana, where we stayed in the worst accommodation we have ever stayed in, and where I also got the pleasure to see Amanda, whom I had not seen in almost two years.

In August of 2021, I visited Amsterdam to celebrate my friend Jorge's birthday for the first time since I had moved to Belgium. In this event I got to see again my friends Satori and Valente. In Amsterdam, Tatiana and I also hung out with my buddy Rodrigo and climbed "Europe's highest swing".

That same month I also had the good fortune that Silvia, one of my best friends since I was 13 years old, was moving to Brussels to do her MBA. She stayed at our place in Leuven until she could find a place of her own, and brought with her the usual laughs and good humor that she usually brings with her. She then found a place in Brussels and we would often see each other either in Brussels, in Leuven, or elsewhere.

For Christmas 2021, Tatiana and I went to Switzerland to spend the holidays with Erika and Nico, including Vale, Pablo, Amélie, Tania, Mathias, and Fran. This

included a very fun trip to the mountains to do some Swiss sledding. We returned to Leuven to receive the New Year in a fun party at Alejo and Marce's place with Enrique, Pancho, Paola, and Furqan.

In February of 2022 I went to Mexico for the first time since I had moved to Belgium. I stayed at my dad's place in my old room. It was great to see all the people that I had not seen in close to two and a half years, and to do all kinds of things. I made tamales with Melissa. I got tacos with my cousin Paulina. I saw my pal Edgar, who I know since we were 7 years old. Pancho and I went to a show of our musician friend, Pedro. I visited my old university and met with Marisela. I met with Saúl, Mario, and Paty to get tacos and coffee. I went to get pulque with Diana, and visited weird architects' houses. I drew some naked people with Diana, Heather, and Joshua. I got tacos with Diana, Heather, and Kim. I got beers and wings with Víctor, Felipe, and Diana Rosales. I got tacos with Tania and Diego, and visited Diego's new apartment. I got tacos and flew drones with Daniel and Pancho. I celebrated my cousin José's birthday and saw Daniela and Maru there. I got tacos with Norma and Miguel at Norma's new house. I went to get *mariscos* with Mireya and Pancho. Melissa dyed my hair. I got tacos with Tania and Rodrigo, and then had coffee with Álvaro. I went to Caro's new apartment. I went to Querétaro with my dad. I saw Clau, Vicente, Ximena, and Lucía over there. I visited Rodrigo and Caro's place in Querétaro. Pancho and I went to Chiapas to visit Chicho. My dad and I met in *La Castellana* with Alex, Fabis, Paola, Alexa, Nico, Erika, Ivette, Ana Karen, Enrique, Maru, José, Daniela, Alfonso, Pablo, and Roberta. I met with the Becerra family in Mayte's house, getting to see my aunts Monis and Gaby, as well as my cousins Fer, Ale, and Andrea. I had tacos with Gerson and Diana. I went for dinner and drinks with Ale, Andrea, and Cris. I went cycling with Miguel and got some pulque (and some tacos again). I got a lot of tacos. I returned to Belgium in late March happy to have spent almost two months in Mexico.

My return from to Belgium came with the good news of a conference in Reykjavík. Iceland had been on my bucket list for a while, so this was a great opportunity for me. In addition to giving a talk there and attending several other very interesting talks, I had the opportunity to go around the island with Álvaro, Ignacio, Gabriel, and Jana. I could also do some cycling over there since I took my road bike with me. That bike ride from Keflavík airport to Reykjavík was certainly something interesting to do.

Around a month later I went to Montréal for another conference. I was happy to go there because I stayed and spent a lot of time with my old friends from university, Omar and Fer. It also allowed me to meet with Karin to get lunch with her, drive around Montréal, and practice my French. I also got to meet some great people at the conference, like Ale, Izzy, Dafne, Irina, Massimilano, Sally, and Antonio.

Some time later, in late summer of 2022, I once again had the good fortune of having two of my best friends, Pancho and Truax, move to Leuven. As my cousin Paulina said, it seemed that I was importing all my friends to Belgium. Having a circle of close friends living in the same city as me has been something I am truly grateful for.

I cannot mention all the events and memories that come to my mind when I write this. Some of the ones that are related to sports include cycling to Mechelen in the freezing cold and rain with Erika and Nico, all for charity. The hikes organized by Jirka and Sanja. Cycling 104 miles with Patrick to visit Baarle-Nassau: the Netherlands within Belgium within the Netherlands. Meeting Sven Nys at a cyclocross event in Overijse with Tatiana, Alex, and Florence. Cycling with José to Pieter and Ioana's place to meet baby Sophia. Doing the Leuven Night Run with David, José, and Clara, with Sara and Nina cheering us in the cold. Cycling 160 km with José, Alex, and Florence in the Paris-Roubaix challenge. Running the Amsterdam Half Marathon with Jorge. Cycling to Brussels with Pancho, Truax, and Carlos. The ultimate frisbee tournament with the people from Kastelpark Arenberg 30, including Remi, Emma, Gerardo, and Ester. Attempting to do the Flandrien Challenge (over 300 Km covering 59 climbs in Belgium) with Patrick and Joan. Going to see one of my friend Bony's Jiu-Jitsu matches in England. Doing lengths in the KU Leuven swimming pool with Tatiana and Pancho.

Others come from traveling, like visiting Luxembourg City with Tatiana, Diana, David, and Pancho. Hanging out in Amsterdam with Satori, Valente, Jorge, and Brenda. Jumping from a bridge into the Zurich river with Diego. Having coffee with Santiago at his place in Amsterdam and him gifting me a box of *Mayordomo* Mexican chocolate. Driving to northern France with Tatiana, Pancho, and Paola just to get a COVID vaccination certificate. Going to Disneyland Paris with Tatiana, Satori, and Valente. Eating Vietnamese food in Paris with Tatiana and Diana. Listening to Christmas carols in Dublin's Christmas market with Diana. Relaxing at the Miami beach with Diana. Visiting Cologne with Tatiana, Paty and Saúl. Eating an unhealthy amount of food in Italy with Saúl, Pancho and Paola. Meeting up with Andrea and Cris in Rome. Partying in Amsterdam with Truax, Diana, Carlos, Victoria, Pancho, Tatiana, Victoria, and Silvia in Truax's birthday. Shooting balls at the golf range with Pancho and Memo in Amsterdam. Driving from Amsterdam to Bruges with Vero, Mariana, and Nacho. Seeing my niece Paola in Amsterdam for an evening. Getting beers with José, Alan, Eric, and Emma in Southampton. Exploring Southampton in the cold and rain with Alan and David. Meeting up with Alan in London. Swimming in the Croatian sea with Diego, Daniel, Jorge, and Luci. Riding in Sander's car to Amsterdam to see Silvia and Diana. Going to the adults-only Christmas market in Hamburg with Edgar, Amanda, Truax and Diana. Racing with sleds in surprisingly steep slopes in the Swiss mountains with Erika, Nico, Gus, Momo, Tatiana, and my dad. And freezing in Lille and Dinant with Tatiana, Alma, and my dad.

Of course, I have many memories from social events like the Thanksgiving dinner at Patrick's house. The office dinner at Anna's place with Saif, Valeria, and Manju. Dinners and Christmas parties at Becca and Alec's. The barbecue at Yannis's place with Yuanyuan, José, Justine, and Jirka. Sofi and Yari preparing brunch at their place in Hasselt for Tatiana and me. The *carnitas asadas* at Enrique's place. The dinner at the Iranian restaurant in Brussels with Babak, Jirka, Sanja, Yannis, Natassa, Pancho, Saúl, and Silvia. Pancho, Diego and I having brunch with Roby in Leuven, and then making Roby hurry on a bike to catch her train. The *noches de dominó* with Pancho, Truax, and Carlos. Toasting for EFFEX's official creation with José, David, Mathias, Ying, Arno, Peter, Bart, and Ivan.

And some random mentions include getting horse steak in Vilvoorde with Silvia and Patrick. Justine helping us move a couch into the office that I shared with Alex and Saif. Patrick and I singing along to *The Devil Went Down to Georgia* over pumpkin pie and getting weird looks from Silvia and Tatiana. David helping me drill some holes on our walls to hang stuff. Going to Bobbejaanland with Rodrigo and Tatiana. Going to Tomorrowland with Tatiana and Sofi, much to my chagrin, but ending up having a lot of fun and randomly running into my university colleagues, Tomás and Arturo. Me getting food poisoning after dinner with Clau, Vicente, Erika, Nico, and Tatiana in Leuven. Visiting museums in Brussels and other places in Belgium with Pancho, Paola, and Tatiana. Many beers, coffees, teas, and cocktails in Leuven's Oude Markt, and countless meals in Ah Quy with so many of the persons that I do and do not mention here.

There are numerous other experiences that I don't bring up in this text, and I'm pretty sure I'm forgetting to mention some persons. But I shared so many events and experiences with so many people in these four years, that I cannot possibly mention and remember all of them. However, I can confidently say that the people that were with me these four years helped immensely in making my PhD journey as fun as possible, and I am truly grateful to have had the chance to live these memories. Special thanks to my work friends Alex, Saif, José and David, who made the DOE group at Kasteelpark Arenberg 30 much more entertaining and enjoyable. And thanks also to my dad for his constant support during my whole academic and professional life. Of course, thanks to Tatiana for deciding to change her whole life and moving with me to make Leuven our new home. And finally, thanks to Erika, without whose support during my bachelor I would most likely not be where I am today, and whose coincidental life changing move to this continent made my own move here much better.

Mario Humberto Becerra Contreras

Leuven, January 2024

Abstract

Many products, services and other items involve mixtures of ingredients, where the mixtures can be expressed as combinations of ingredient proportions, for instance, flour, water and yeast to make bread. In many cases, the result of the mixture may also depend on the way in which the mixtures are processed, such as the baking time and baking temperature for a bread. These types of variables are generally called process variables.

In mixture experiments with or without process variables, the researchers are generally interested in one or more characteristics of the mixture. In this thesis, the characteristic of interest is the preference of respondents. A common approach to study the preference of respondents is through choice modeling and discrete choice experiments.

Discrete choice experiments are frequently used to quantify consumer preferences by having respondents choose between different alternatives. Choice experiments involving mixtures of ingredients and process variables have been largely overlooked in the literature, even though many products, services and items can be described as mixtures of ingredients. As a consequence, little research has been done on the analysis and optimal design of choice experiments involving mixtures and process variables.

In experiments with mixtures, it is crucial to obtain models that yield precise predictions for any combination of ingredient proportions and process variables. This is because the goal of these types of experiments generally is to find the combination of ingredient proportions and process variables that optimizes the respondents' utility. Because it focuses on minimizing the average prediction variance, the I-optimality criterion is the most suitable for designing choice experiments with mixtures and process variables.

In this thesis, we show how to combine mixture experiments and discrete choice experiments to quantify and model preferences for choice options that can be viewed as mixtures of ingredients. We construct Bayesian I-optimal designs, compare them with their Bayesian D-optimal counterparts, and show that the former designs perform substantially better than the latter in terms of the variance of the predicted utility.

We applied these concepts in a series of choice experiments involving a particular species of fruit fly, called *Drosophila suzukii*, that has invaded Europe and has

been causing damage to local agriculture. Because of the damage that this fly causes, methods that can trap and monitor these fruit flies are sought. Since the flies use visual and olfactory cues to find food, most traps are based on attraction using color or smell. Hence, studying the color preference of *Drosophila suzukii* is useful for the creation of traps.

We designed and carried out several choice experiments with *Drosophila suzukii* in order to model their color preferences. We built a mixture-amount model involving individual LED colors and their intensity to model the attractiveness of colors and total intensity, and embedded that model in a mixed logit choice model. The mixture-amount model is a particular case of a mixture-process variable model where the amount acts as the single process variable. Using our model, we identified the most attractive combinations of colors, and we ran a confirmatory experiment to test the attractiveness of these colors. We found that *Drosophila suzukii* are attracted to UV light and that a combination of UV light and other colors yields higher utility. We conclude that our methodology was appropriate to identify the most attractive colors for the *Drosophila suzukii* and can be used to create traps.

Beknopte samenvatting

Veel producten en diensten kunnen beschouwd worden als mengsels van ingrediënten. Mengsels kunnen altijd uitgedrukt worden als combinaties van de proporties van de ingrediënten, bijvoorbeeld bloem, water en gist om brood te maken. In veel gevallen hangen de uiteindelijke eigenschappen van een mengsel ook af van de manier waarop het mengsel wordt verwerkt (denk aan de baktijd en de baktemperatuur van een brood). De variabelen die de verwerking van een mengsel karakteriseren worden procesvariabelen genoemd.

Bij mengselexperimenten met of zonder procesvariabelen zijn de onderzoekers doorgaans geïnteresseerd een model dat de eigenschappen van het mengsel beschrijft. In dit proefschrift is de bestudeerde eigenschap de voorkeur van respondenten. Een gebruikelijke werkwijze om de voorkeur van respondenten te bestuderen, te kwantificeren en te modelleren is door middel van discrete keuze-experimenten en keuzemodellen.

Discrete keuze-experimenten worden vaak gebruikt om de voorkeuren van consumenten te kwantificeren door respondenten herhaaldelijk te laten kiezen tussen verschillende alternatieven. Keuze-experimenten met mengsels van ingrediënten en procesvariabelen werden in de literatuur grotendeels over het hoofd gezien, ook al kunnen veel producten en diensten worden omschreven als mengsels van ingrediënten. Bijgevolg werd tot nog toe weinig onderzoek gedaan naar de analyse en het optimale ontwerp van keuze-experimenten met mengsels en procesvariabelen.

Bij experimenten met mengsels is het van cruciaal belang om modellen te verkrijgen die nauwkeurige voorspellingen opleveren voor elke combinatie van ingrediëntenproporties en procesvariabelen. Dit komt doordat het doel van dit soort experimenten over het algemeen is om de combinatie van ingrediëntenverhoudingen en procesvariabelen te vinden die het gepercipieerde nut en dus de voorkeur van de respondenten optimaliseert. Omdat het zich richt op het minimaliseren van de gemiddelde voorspellingsvariantie is het I-optimaliteitscriterium het meest geschikt voor het ontwerpen van keuze-experimenten met mengsels en procesvariabelen.

In dit proefschrift laten we zien hoe experimenten met mengsels en procesvariabelen en discrete-keuze experimenten kunnen worden gecombineerd om de voorkeuren van consumenten voor alternatieven die als mengsels kunnen worden beschouwd, te kwantificeren en te modelleren. We construeren Bayesiaanse I-optimale ontwerpen,

vergelijken ze met hun Bayesiaanse D-optimale tegenhangers, en laten zien dat de eerstgenoemde ontwerpen substantieel beter presteren dan de laatstgenoemde in termen van de variantie van het voorspelde nut.

We hebben de nieuw ontwikkelde concepten toegepast in een reeks keuze-experimenten waarbij een bepaalde soort fruitvlieg betrokken was, genaamd *Drosophila suzukii*, die Europa is binnengedrongen en schade heeft aangericht aan de lokale landbouw. Vanwege de schade die deze vlieg aanricht, wordt er gezocht naar methoden om deze fruitvliegen te vangen en te monitoren. Omdat de vliegen visuele en reuksignalen gebruiken om voedsel te vinden, zijn de meeste vallen gebaseerd op aantrekking met behulp van kleur of geur. Daarom is het bestuderen van de kleurvoorkeur van *Drosophila suzukii* nuttig voor het maken van vallen.

Voor onze keuze-experimenten met *Drosophila suzukii* hebben we een mengselhoeveelheidsmodel gebruikt waarin de individuele LED-kleuren fungeerden als ingrediënten van het kleurmengsel en de totale intensiteit van het licht als hoeveelheidsvariabele dienst deed. Dat mengselhoeveelheidsmodel hebben we vervolgens ingebed in een keuzemodel. Het mengsel-hoeveelheidsmodel is een speciaal geval van een model voor mengsels en procesvariabelen, waarbij de hoeveelheid fungert als de enige procesvariabele. Met behulp van ons uiteindelijke model hebben we de aantrekkelijkste kleurencombinaties geïdentificeerd en hebben we een confirmatorisch experiment uitgevoerd om de attractiviteit van deze kleuren te verifiëren. We ontdekten dat *Drosophila suzukii* zich aangetrokken voelen tot UV-licht, maar dat een combinatie van UV-licht en andere kleuren een groter nut oplevert. We concluderen dat onze methodologie geschikt was om de meest aantrekkelijke kleuren voor de *Drosophila suzukii* te identificeren en dat de methodologie nuttig is om geschikte vallen te creëren.

Contents

Abstract	ix
Beknopte samenvatting	xi
Contents	xiii
1 Introduction	1
2 Bayesian I-optimal designs for choice experiments with mixtures	5
2.1 Introduction	6
2.2 Models	7
2.2.1 Models for data from mixture experiments	7
2.2.2 Multinomial logit model for choice data	8
2.2.3 Model for choice data concerning mixtures	9
2.3 Optimal design criteria	10
2.3.1 Information matrix	10
2.3.2 D-optimal designs	11
2.3.3 I-optimal designs	12
2.4 Construction of I-optimal designs	14
2.5 Results	16
2.5.1 Cocktail preferences	16
2.5.2 Artificial sweetener experiment	19
2.6 Discussion	21
3 Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables	24
3.1 Introduction	25
3.2 Models	26
3.2.1 Models for data from mixture experiments including process variables	26
3.2.2 Multinomial logit model for choice data	28
3.2.3 Model for choice data concerning mixtures and process variables	29
3.3 Optimal design criteria	30
3.3.1 Information matrix	31
3.3.2 D-optimal designs	31

3.3.3	I-optimal designs	32
3.3.4	Numerical approximation to optimality criteria	35
3.3.5	Construction of D- and I-optimal designs	35
3.4	Results	36
3.4.1	Cocktail example	37
3.4.2	Fish patty example	40
3.5	Discussion	42
4	Finding fruit flies' favorite color: a discrete choice experiment	46
4.1	Introduction	47
4.2	Models	49
4.2.1	Models for data from mixture-amount experiments	50
4.2.2	Multinomial logit model for choice data	51
4.2.3	Model for choice data concerning mixture-amount experiments	51
4.2.4	Mixed logit model	52
4.2.5	No-choice option	54
4.2.6	Estimation	55
4.3	I-optimal designs	56
4.4	Experimental setup	59
4.5	Results	60
4.5.1	Bayesian estimation	61
4.5.2	Model validation	66
4.5.3	Model predictions	67
4.6	Discussion	70
5	Conclusions and future work	74
5.1	Conclusions	74
5.2	Future work	75
A	Appendix	77
A.1	Design tables of artificial sweetener experiment	77
A.2	Design tables of cocktail experiment	82
Curriculum vitae		93
		95

Chapter 1

Introduction

Many products, services, and other items can be described as mixtures of ingredients. For example, the chemicals that are used to create a pesticide (Cornell, 2011); cement, water, and sand to make concrete (Cornell, 2002); media used in advertising campaigns (Goos et al., 2019); components of a mobility budget (Zijlstra et al., 2019); the wheat varieties used to bake bread (Rehman et al., 2007); and the ingredients used to make a drink such as mango juice, lime juice, and blackcurrant syrup (Courcoux & Séménou, 1997b; Goos & Hamidouche, 2019). In mixture experiments, the products, services, or items under investigation are expressed as combinations of proportions of ingredients and the researchers' interest is generally in one or more characteristics of the mixture. In this work, the characteristic of interest is the preference of respondents.

A common approach to study the preference of respondents is through choice modeling and discrete choice experiments. This approach has been successfully applied in marketing (Rossi et al., 2012; Train, 2009), transportation (Khademi & Timmermans, 2012; Khademi et al., 2013; Rossetti et al., 2023; Zijlstra et al., 2019), ecology (Melero et al., 2018; Vardakis et al., 2015), environmental economics (Bennett & Blamey, 2001; Torres et al., 2013; Vojáček, Pecáková, et al., 2010), sports (Balliauw et al., 2020; Becerra et al., 2023), and food research (Apellado-Buenaventura & Valmorida, 2021; Lizin et al., 2022; Skaltsi et al., 2022).

Discrete choice experiments are carried out by presenting respondents sets of alternatives, called *choice sets*, and letting them choose between the alternatives. The respondents repeat this task several times with different choice sets. One decision that an experimenter has to make is how many alternatives will appear in each choice set. It is common to have the respondents do pairwise comparisons with choice sets that involve two alternatives each. The modeling of paired comparisons has existed for almost 100 years, being first introduced in 1927 by Thurstone (1927). In 1952, the Bradley-Terry model was introduced, which is a logit model for paired evaluations and can be used for ranking competitors in sports and other competitions (Agresti, 2012; Bradley & Terry, 1952). David (1963) mentioned that paired comparisons at the time were often used in preference testing and

modeling. The suggestion of doing paired comparisons is due to the fact that performing pairwise comparisons is easier for respondents than comparing bigger sets of alternatives (Courcoux & Sémenou, 1997a; Ruseckaite et al., 2017).

In addition to the number of alternatives per choice set, another decision to be made in the design of choice experiments is the size of the experiment. This, of course, depends on many things, such as budgetary constraints, the objectives of the experiment, and the complexity of the phenomenon that is being studied. Assele et al. (2023) review a series of different approaches to calculate the required sample size in a discrete choice experiment. These approaches range from rules of thumb, such as a minimum of 100 or 300 respondents, to power calculations taking into account the value of the parameters. The authors show how to use a regression-based approach to calculate sample size under certain particular assumptions.

In the literature, the size of the experiments varies considerably. For example, Courcoux and Sémenou (1997b) had 60 respondents each of whom performed eight pairwise comparisons, Zijlstra et al. (2019) had 817 respondents each of whom performed 16 pairwise comparisons, Becerra et al. (2023) had 589 respondents each of whom performed ten pairwise comparisons, Rossetti et al. (2023) had 2315 respondents each of whom performed 8 pairwise comparisons, and Boonaert et al. (2023) had 426 respondents each of whom performed 12 pairwise comparisons. In practice, it is also common to add certain control questions to make sure that respondents are paying attention. One example is adding a couple of choice sets where the better choice is extremely obvious. Another example is repeating the same choice set twice with reversed order to make sure respondents are being consistent in their decisions. In our work, we assume that the experimenter has fixed the number of choice sets and the number of alternatives per choice set.

Despite the fact that numerous products, services, and other items involve mixtures of ingredients, literature concerning choice experiments with mixtures is scarce. Some of the few works on this topic include the work by Courcoux and Sémenou (1997b), who performed a taste experiment to model the preferences for cocktails involving different proportions of mango juice, lime juice, and blackcurrant syrup. Goos and Hamidouche (2019) used the data from Courcoux and Sémenou (1997b) and explained how to embed Scheffé models for data from mixture experiments in logit type models for choice experiments. Khademi and Timmermans (2012) analyzed the response of travelers to mixtures of transport charges subject to budget constraints. Khademi et al. (2013) discussed a choice experiment involving a mixture of road toll, congestion pricing, and parking price. Yang et al. (2016) used a mixture choice experiment to measure context-dependent responses to accumulative energy charges under budget constraints. Zijlstra et al. (2019) performed a choice experiment to investigate the preferred mobility budget mixture for employees of a company. Boonaert et al. (2023) used a choice experiment concerning the desired composition of a family in sub-Saharan Africa, where the family composition was considered a mixture of boys and girls with different education levels.

Experiments, including choice experiments, are expensive, cumbersome and time-consuming. Therefore, efficient experimental designs are required so that the experiments provide reliable information, be it for statistical modeling, for precise

estimation of model parameters, or for precise predictions. Optimal design of experiments is the branch of statistics dealing with the construction of efficient experimental designs.

There is little literature about the optimal design of choice experiments with mixtures. Ruseckaite et al. (2017) compared two algorithms to find D-optimal designs for choice experiments with mixtures. D-optimality is arguably the most commonly used metric for optimal designs. A D-optimal design approach can be viewed as an estimation-based approach, because it is intended to minimize the generalized variance of the estimators of the model parameters. D-optimal experimental designs are thus a good choice if obtaining low-variance estimators is the main goal of the choice experiment. However, in experiments with mixtures, the goal generally is to optimize the composition of the mixture to maximize consumer preference. Therefore, in mixture experiments, it is crucial to obtain models that yield precise predictions for any combination of ingredient proportions.

Two potential criteria for this are I- and G-optimality, both of which focus on the prediction variance and should therefore allow a better identification of the mixture that maximizes consumer preference. I-optimal designs minimize the average prediction variance over the experimental region, whereas G-optimal designs minimize the maximum prediction variance over the experimental region. Rodriguez et al. (2010) studied G-optimality and argued that in order to minimize the maximum variance of the prediction, the experimenter may have to accept worse prediction variances over most of the region of interest, which might not be acceptable in many cases. This same reason inspired Goos et al. (2016) to focus on I-optimality instead of G-optimality in their work about the designs involving mixtures of ingredients. In this thesis, we follow this line of thinking and therefore also focus on I-optimal designs.

The main goal of this thesis is to lay out the methodology required to design D- and I-optimal choice experiments involving mixtures of ingredients and to analyze the resulting data. We introduce an optimization algorithm that can be used to create optimal designs for choice experiments with mixtures. We start with a model for mixtures that deals only with proportions of ingredients, and then expand it to include variables that are unrelated to the mixture's composition. These additional variables are generally referred to as *process variables*. We then use the optimization algorithm to create Bayesian I-optimal designs for a choice experiment that intends to model the color preferences of fruit flies. We finally analyze the results from this experiment.

The outline of the thesis is as follows. In Chapter 2, we introduce a computationally efficient definition for I-optimal designs for choice experiments involving mixtures of ingredients and show how to embed the new I-optimality criterion in a coordinate-exchange algorithm for constructing D- and I-optimal designs. We use the same type of Scheffé models as Goos and Hamidouche (2019) and we revisit two examples from the literature to study the difference between D- and I-optimal designs.

In Chapter 3, we extend the models and algorithms from Chapter 2 to deal with situations where the preference for mixtures of ingredients also depends on characteristics unrelated to the composition of the mixture. For example, the

preference for a cocktail might depend on the temperature at which it is served, or the firmness of a fish patty depends not only on the types of fish used, but also on baking temperature, baking time, and frying time (Cornell, 1988, 2002; Cornell & Gorman, 1984; Goos, 2022). The color, aroma, taste, texture and mouthfeel of *pastillas de leche*, a popular Filipino candy, depend on baking time and temperature in addition to the proportions of the mixture ingredients such as cornstarch, flour, glucose, sugar and milk (Apellado-Buenaventura & Valmorida, 2021). The aroma, hardness, crispness, color and fracture force of apple biscuits are affected by the mixture ingredients and the microwave blanching of the apples (Skaltsi et al., 2022). In the aforementioned works by Khademi et al. (2013), Yang et al. (2016), Zijlstra et al. (2019), and Boonaert et al. (2023), there was a variable related to the total amount of the mixture in their models. In the general literature on mixture experiments, variables such as baking temperature, baking time, frying time, serving temperature, microwave blanching, and total amount of the mixture are typically called *process variables* (Goos & Jones, 2011). In Chapter 3, we elaborate on how to embed process variables in the modeling and construction of D- and I-optimal designs of choice experiments involving mixtures.

Finally, in Chapter 4, we use the models from Chapter 3 to study the color preferences of the species of fruit fly called *Drosophila suzukii*. This species has invaded Europe and has been causing damage to local agriculture. Because of that damage, methods to trap and monitor these fruit flies are sought. Since the flies use visual and olfactory cues to find food, most traps are based on attraction using color or smell. Hence, studying the color preference of *Drosophila suzukii* is useful for the creation of traps.

We designed and carried out several choice experiments with *Drosophila suzukii* in order to model their color preferences and identify the most attractive colors. We built a mixture-amount model of individual LED colors and their total intensity to model the attractiveness of colors, and embedded that model in a mixed logit choice model. The mixture-amount model is a particular case of a mixture-process variable model. Using our model, we identified the most attractive combinations of colors, and we ran a confirmatory experiment to test the attractiveness of these colors. We found that *Drosophila suzukii* are attracted to UV light and that a combination of UV light and other colors is even more attractive than pure UV light. We conclude that our methodology was appropriate to identify the most attractive colors for the *Drosophila suzukii* and can be used to create traps.

Chapter 2

Bayesian I-optimal designs for choice experiments with mixtures¹

Abstract

Discrete choice experiments are frequently used to quantify consumer preferences by having respondents choose between different alternatives. Choice experiments involving mixtures of ingredients have been largely overlooked in the literature, even though many products and services can be described as mixtures of ingredients. As a consequence, little research has been done on the optimal design of choice experiments involving mixtures. The only existing research has focused on D-optimal designs, which means that an estimation-based approach was adopted. However, in experiments with mixtures, it is crucial to obtain models that yield precise predictions for any combination of ingredient proportions. This is because the goal of mixture experiments generally is to find the mixture that optimizes the respondents' utility. As a result, the I-optimality criterion is more suitable for designing choice experiments with mixtures than the D-optimality criterion because the I-optimality criterion focuses on getting precise predictions with the estimated statistical model. In this chapter, we study Bayesian I-optimal designs, compare them with their Bayesian D-optimal counterparts, and show that the former designs perform substantially better than the latter in terms of the variance of the predicted utility.

¹This chapter is, with minor adaptations, a copy of the following article: Becerra, Mario & Goos, Peter (2021) *Bayesian I-optimal designs for choice experiments with mixtures*, Chemometrics and Intelligent Laboratory Systems, 217, 104395, DOI: 10.1016/j.chemolab.2021.104395

2.1 Introduction

Discrete choice experiments are frequently used to quantify consumer preferences. These experiments collect stated preference data and are carried out by presenting respondents sets of alternatives, called choice sets, and having them choose between the alternatives. The respondents repeat this task several times with different choice sets. It is common to have the respondents do pairwise comparisons with choice sets that involve two alternatives each.

Discrete choice experiments have been successfully applied in areas such as marketing (Rossi et al., 2012; Train, 2009), transportation (Zijlstra et al., 2019), health care (Luyten et al., 2015), ecology (Fletcher Jr et al., 2015; Melero et al., 2018; Vardakakis et al., 2015), and environmental economics (Bennett & Blamey, 2001; Torres et al., 2013; Vojáček, Pecáková, et al., 2010). However, choice experiments involving mixtures have been largely overlooked in the literature, even though many products and services can be described as mixtures of ingredients.

Examples of mixture ingredients include the chemicals that are used to create a pesticide (Cornell, 2011); media used in advertising campaigns (Goos et al., 2019); components of a mobility budget such as car with fuel card and public transport card (Zijlstra et al., 2019); cement, water, and sand to make concrete (Cornell, 2002); the wheat varieties used to bake bread (Rehman et al., 2007); and the ingredients used to make a drink such as mango juice, lime juice, and blackcurrant syrup (Courcoux & Séménou, 1997b; Goos & Hamidouche, 2019). In mixture experiments, the products and services under investigation are expressed as combinations of proportions of ingredients and the researchers' interest is generally in one or more characteristics of the mixture. In this chapter, the characteristic of interest is the preference of respondents. Choice experiments are ideal to collect data for quantifying and modeling preferences for mixtures.

The first example of a discrete choice experiment concerning mixtures was published by Courcoux and Séménou (1997b). The experiment is actually a taste experiment to model the preferences for cocktails involving different proportions of mango juice, lime juice, and blackcurrant syrup. The resulting experimental data involve the responses of sixty people, each making eight pairwise comparisons of different cocktails. Thus, eight times in total, each respondent had to taste two different cocktails and say which one they preferred. This means that the choice experiment comprised $60 \times 8 = 480$ choice sets, each of size 2. Goos and Hamidouche (2019) described how to combine Scheffé models for data from mixture experiments with the logit type models typically used for choice experiments, and demonstrated the usefulness of the resulting combined models using the data from Courcoux and Séménou (1997b).

As discrete choice experiments in general, and taste experiments in the form of a discrete choice experiment in particular, are expensive, cumbersome and time-consuming, efficient experimental designs are required so that the experiments provide reliable information for statistical modeling, precise estimation of model parameters and precise predictions. Optimal design of experiments is the branch of statistics dealing with the construction of efficient experimental designs.

Little research has been done concerning the optimal design of choice experiments with mixtures. For their experiment, Courcoux and Sémenou (1997b) used an elegant ad hoc design construction combining a simplex centroid design to define the mixtures used and a balanced incomplete block design to define subsets of the mixtures. Assuming a multinomial logit model, Ruseckaite et al. (2017) compared two algorithms to find D-optimal designs for choice experiments with mixtures. A D-optimal design is an experimental design that maximizes the determinant of the Fisher information matrix. The D-optimal design approach can be viewed as an estimation-based approach, because it is intended to minimize the generalized variance of the estimators of the model parameters. D-optimal experimental designs are thus a good choice if obtaining low-variance estimators is the main goal of the choice experiment. However, in experiments with mixtures, the goal generally is to optimize the composition of the mixture to maximize consumer preference. Therefore, in mixture experiments, it is crucial to obtain models that yield precise predictions for any combination of ingredient proportions. As a result, I-optimal experimental designs are more suitable for choice experiments with mixtures than D-optimal ones. I-optimal designs minimize the average prediction variance and therefore allow a better identification of the mixture that maximizes the consumer preference. In this chapter, we therefore study I-optimal designs for discrete choice experiments involving mixtures.

The rest of the chapter is organized as follows. In Section 2.2, we describe how to incorporate mixtures into the multinomial logit choice model. In Section 2.3, we describe the D- and I-optimality criteria. In Section 2.4, we outline the design construction algorithm we use. In Section 2.5, we present our computational results. Finally, in Section 2.6, we summarize and discuss our work.

2.2 Models

In this section, we introduce the most commonly used models for data from mixture experiments as well as the multinomial logit model for choice data, and explain how to combine the two models for data from choice experiments involving mixtures.

2.2.1 Models for data from mixture experiments

Mixture experiments involve two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture. Each mixture is described as a combination of q ingredient proportions, with the constraint that these proportions sum up to one. Due to this constraint, a classical regression model involving an intercept and linear terms in the ingredient proportions exhibits perfect collinearity. Therefore, researchers must use dedicated regression models when analyzing data from mixture experiments. The most commonly used family of models for data from mixture experiments is the Scheffé family (Scheffé, 1958, 1963). The most popular Scheffé models are the first-order, second-order, and special-cubic models.

Denoting the response in a traditional mixture experiment with a continuous outcome by Y and the q ingredient proportions by x_1, x_2, \dots, x_q , with $x_i \geq 0$ and $\sum_{i=1}^q x_i = 1$, the first-order Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \varepsilon.$$

The second-order Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \varepsilon,$$

and, finally, the special-cubic Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_i x_k x_l + \varepsilon.$$

In all three cases, ε denotes the error term, which, traditionally, is assumed to be normally distributed.

2.2.2 Multinomial logit model for choice data

The multinomial logit model builds on random-utility theory and assumes that a respondent in a choice experiment faces S choice sets involving J alternatives. The model assumes that, within each choice set $s \in \{1, \dots, S\}$, each respondent chooses the alternative that has the highest perceived utility. Therefore, the probability that a respondent chooses alternative $j \in \{1, \dots, J\}$ in choice set s , denoted by p_{js} , is the probability that the perceived utility of alternative j in choice set s , denoted by U_{js} , is larger than that of the other alternatives in the choice set:

$$p_{js} = \mathbb{P}[U_{js} > \max(U_{1s}, \dots, U_{j-1,s}, U_{j+1,s}, \dots, U_{Js})].$$

Since each alternative in a choice set has a set of observable attributes that characterize it, the perceived utility U_{js} can be expressed as

$$U_{js} = \mathbf{f}^T(\mathbf{x}_{js})\boldsymbol{\beta} + \varepsilon_{js}, \quad (2.1)$$

where \mathbf{x}_{js} is the vector that contains the attributes corresponding to alternative j in choice set s , $\mathbf{f}(\mathbf{x}_{js})$ represents the model expansion of this attribute vector, and $\boldsymbol{\beta}$ is the vector containing the model parameters. The model parameters contained within $\boldsymbol{\beta}$ express the preferences of the respondents for the alternatives' attributes. The error terms ε_{js} in Equation (2.1) are assumed to be independent and identically Gumbel distributed. The Gumbel distribution is also known as the generalized extreme value distribution type I and as the log-Weibull distribution. As a result of the distributional assumption, it can be shown that

$$p_{js} = \frac{\exp[\mathbf{f}^T(\mathbf{x}_{js})\boldsymbol{\beta}]}{\sum_{t=1}^J \exp[\mathbf{f}^T(\mathbf{x}_{ts})\boldsymbol{\beta}]},$$

which is called the softmax function in some research areas.

2.2.3 Model for choice data concerning mixtures

In this chapter, where we focus on choice experiments involving mixtures, we assume that the attributes of the alternatives in a choice experiment are the ingredients of a mixture. Consequently, we assume that the attribute vector \mathbf{x}_{js} contains the q ingredient proportions x_1, x_2, \dots, x_q and that $\mathbf{f}(\mathbf{x}_{js})$ represents the model expansion of these proportions. Following Ruseckaite et al. (2017), we base the polynomial expansion $\mathbf{f}(\mathbf{x}_{js})$ on the first-order, second-order, or special-cubic Scheffé model; depending on the complexity of the respondents' preferences. To explain how this is done, we start from the special-cubic model. The derivation of the models based on the first-order and second-order Scheffé model is analogous.

When starting from the special-cubic Scheffé model, the most natural thing to do would be to write the perceived utility U_{js} of a mixture alternative j in choice set s as

$$U_{js} = \sum_{i=1}^q \beta_i x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js},$$

with the error terms ε_{js} assumed to be independent and identically Gumbel distributed. However, due to the constraint that the ingredient proportions must sum up to one, this leads to an inestimable multinomial logit model. As a matter of fact, due to that constraint, we can rewrite x_{qjs} as $1 - x_{1js} - \dots - x_{q-1,js}$ and U_{js} as

$$\begin{aligned} U_{js} &= \sum_{i=1}^{q-1} \beta_i x_{ijs} + \beta_q (1 - x_{1js} - \dots - x_{q-1,js}) + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js} \\ &= \beta_q + \sum_{i=1}^{q-1} (\beta_i - \beta_q) x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} + \varepsilon_{js}. \end{aligned} \tag{2.2}$$

This final expression for the perceived utility U_{js} involves a constant, β_q . Because the multinomial logit model only takes into account differences in utility, that constant causes any choice model based on U_{js} to be ill-defined and therefore inestimable. This can be easily remedied by dropping β_q and using the following

expression for the perceived utility:

$$\begin{aligned}
U_{js} &= \mathbf{f}^T(\mathbf{x}_{js})\boldsymbol{\beta} + \varepsilon_{js} \\
&= \sum_{i=1}^{q-1} \beta_i^* x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_{ijs} x_{kjs} + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_{ijs} x_{kjs} x_{ljs} \quad (2.3) \\
&\quad + \varepsilon_{js},
\end{aligned}$$

with $\beta_i^* = \beta_i - \beta_q$ for $i \in \{1, \dots, q-1\}$. In this expression,

$$\mathbf{x}_{js} = (x_{1js}, x_{2js}, \dots, x_{qjs})^T$$

and

$$\begin{aligned}
\mathbf{f}(\mathbf{x}_{js}) &= (x_{1js}, x_{2js}, \dots, x_{q-1,js}, x_{1js}x_{2js}, \dots, x_{q-1,js}x_{qjs}, \\
&\quad x_{1js}x_{2js}x_{3js}, \dots, x_{q-2,js}x_{q-1,js}x_{qjs})^T.
\end{aligned}$$

The parameter vector $\boldsymbol{\beta}$ for the special-cubic model is then given by

$$\boldsymbol{\beta} = (\beta_1^*, \beta_2^*, \dots, \beta_{q-1}^*, \beta_{1,2}, \dots, \beta_{q-1,q}, \beta_{123}, \dots, \beta_{q-2,q-1,q})^T.$$

This vector has $r = (q^3 + 5q)/6 - 1$ elements in the event the special-cubic model is used. For the first- and second-order Scheffé models, the parameter vector $\boldsymbol{\beta}$ involves $r = q - 1$ and $r = (q^2 + q)/2 - 1$ elements, respectively, after implementing the remedy to make the model estimable.

2.3 Optimal design criteria

2.3.1 Information matrix

To select D- and I-optimal experimental designs, it is required to know the information matrix of the model. For the multinomial logit model, the information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})$ is obtained as the sum of the information matrices of each of the S choice sets (Kessels et al., 2006):

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\beta}) = \sum_{s=1}^S \mathbf{X}_s^T (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T) \mathbf{X}_s,$$

with $\mathbf{p}_s = (p_{1s}, \dots, p_{Js})^T$, $\mathbf{P}_s = \text{diag}(\mathbf{p}_s)$, $\mathbf{X}_s^T = [\mathbf{f}(\mathbf{x}_{js})]_{j \in \{1, \dots, J\}}$ denoting the model matrix corresponding to all alternatives in choice set s , and $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_S]$ denoting the model matrix for all choice sets. The information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\beta})$ is of dimension $r \times r$. The inverse of the information matrix is the asymptotic variance-covariance matrix of the parameter estimates.

Note that the information matrix depends on the unknown parameter vector β , through the choice probabilities contained within \mathbf{p}_s and \mathbf{P}_s . This is typical for models that are not linear in the parameters, such as discrete choice models, and it implies that prior information is needed to find optimal designs (Atkinson & Haines, 1996; Kessels et al., 2006; Ruseckaite et al., 2017). The prior information can be in the form of a point estimate, or in the form of a prior distribution. An optimal design that uses only a point estimate is called a locally optimal design, whilst one that uses a prior distribution is called a Bayesian optimal design.

In optimal experimental design, there are different criteria to measure the quality of a design and the corresponding model matrix \mathbf{X} . As pointed out in Goos and Jones (2011), the two most widely used criteria for selecting experimental designs in business and industry are D-optimality and I-optimality (also called V-optimality; see Goos and Syafitri (2014)). The former is an estimation-oriented criterion because it focuses on a precise model estimation, while the latter is a prediction-oriented criterion because it focuses on getting precise predictions with the estimated statistical model.

2.3.2 D-optimal designs

The D-optimality criterion is the most traditional metric used in the design of choice experiments (Bliemer & Rose, 2010, 2011; Bliemer et al., 2009; Burgess & Street, 2005; Grasshoff et al., 2003; Kessels et al., 2011). For a model matrix \mathbf{X} and parameter vector β , the D-optimality criterion can be defined as

$$\mathcal{D} = \det ([\mathbf{I}^{-1}(\mathbf{X}, \beta)])^{\frac{1}{r}}, \quad (2.4)$$

where $\mathbf{I}^{-1}(\mathbf{X}, \beta)$ is the asymptotic variance-covariance matrix and the inverse of the information matrix of the parameter estimates. A design that minimizes Equation (2.4) using a point estimate of β is called a locally D-optimal design. The problem with locally D-optimal designs is that they may perform poorly for values of the parameter vector β for which they were not optimized. This weakness is, of course, highly relevant because the true values of the model parameters are unknown.

Bayesian designs take into account prior information and uncertainty about the parameter vector β . More specifically, they are based on a prior distribution $\pi(\beta)$ which summarizes the prior knowledge concerning β . Most of the Bayesian design constructions in the choice experiments literature adopt the approach where the D-optimality criterion is averaged over the prior distribution (Bliemer & Rose, 2011; Bliemer et al., 2009; Kessels et al., 2011). This is also the approach we use. Therefore, we define the Bayesian D-optimality criterion for the multinomial logit model as

$$\mathcal{D}_B = \int_{\mathbb{R}^r} [\det (\mathbf{I}^{-1}(\mathbf{X}, \beta))]^{\frac{1}{r}} \pi(\beta) d\beta, \quad (2.5)$$

where $\pi(\beta)$ is the prior distribution of β . We refer to a design that minimizes the Bayesian D-optimality criterion as a Bayesian D-optimal design, even though the criterion does not take into account the posterior distribution and some authors therefore prefer to call these designs pseudo-Bayesian designs (Ryan et al., 2016).

2.3.3 I-optimal designs

The I-optimality criterion is generally defined as the average of the prediction variance over the experimental region, which we denote by χ and which is the $(q-1)$ -dimensional simplex if there are no constraints on the ingredient proportions other than those mentioned in Section 2.2.1. Now, when using choice models, there are two ways in which we can define I-optimality. If the goal is to predict choice probabilities, the I-optimality criterion is the average variance of the predicted choice probabilities. If the goal is to predict perceived utilities, the I-optimality criterion is the average variance of the predicted utilities.

I-optimality for predicted choice probabilities

Kessels et al. (2009) computed I-optimal designs based on the variance of the predicted choice probability, $\text{Var}[\hat{p}_{js}]$. Since this prediction variance cannot be calculated analytically, they approximated it using a first-order Taylor series expansion of the choice probability:

$$\text{Var}[\hat{p}_{js}] \approx \mathbf{c}^T(\mathbf{x}_{js}) \mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}) \mathbf{c}(\mathbf{x}_{js}),$$

where \hat{p}_{js} denotes the predicted choice probability, and

$$\mathbf{c}(\mathbf{x}_{js}) = \frac{\partial p_{js}}{\partial \boldsymbol{\beta}} = p_{js} \left(\mathbf{x}_{js} - \sum_{t=1}^J p_{ts} \mathbf{x}_{ts} \right).$$

As a consequence, the I-optimality criterion of Kessels et al. (2009) is

$$\begin{aligned} \frac{\int_{\chi} \text{Var}[\hat{p}_{js}] d\mathbf{x}_{js}}{\int_{\chi} d\mathbf{x}_{js}} &= \frac{\int_{\chi} \mathbf{c}^T(\mathbf{x}_{js}) \mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}) \mathbf{c}(\mathbf{x}_{js}) d\mathbf{x}_{js}}{\int_{\chi} d\mathbf{x}_{js}} \\ &= \frac{\int_{\chi} \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}) \mathbf{c}(\mathbf{x}_{js}) \mathbf{c}^T(\mathbf{x}_{js})] d\mathbf{x}_{js}}{\int_{\chi} d\mathbf{x}_{js}} = \frac{\text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta}) \mathbf{W}_p(\boldsymbol{\beta})]}{\int_{\chi} d\mathbf{x}_{js}}, \end{aligned} \quad (2.6)$$

where

$$\mathbf{W}_p(\boldsymbol{\beta}) = \int_{\chi} \mathbf{c}(\mathbf{x}_{js}) \mathbf{c}^T(\mathbf{x}_{js}) d\mathbf{x}_{js} \quad (2.7)$$

is referred to as the moments matrix in the literature on I-optimality and the subscript p refers to the fact that this is the moments matrix for the I-optimality criterion based on choice probabilities. The denominator in Equation (2.6) is the volume of the experimental region χ . This denominator can be safely ignored when constructing I-optimal designs because it is constant for all designs for a given experiment.

Due to the fact that Kessels et al. (2009) considered only categorical attributes in their choice experiments, they were able to calculate the moments matrix exactly for any given parameter vector $\boldsymbol{\beta}$. This is impossible when continuous attributes are considered, because there is no closed-form solution for the integral in Equation

(2.7). Since the attributes we consider in this chapter are ingredient proportions, we also cannot calculate $\mathbf{W}_p(\beta)$ exactly. The solution to this problem would be to numerically approximate the integral in Equation (2.7).

In our view, basing the I-optimality criterion for choice experiments with mixtures on predicted probabilities suffers from four weaknesses. First, there is no analytical expression for the variance of a predicted choice probability, which necessitates a first approximation. Second, there is no closed form expression for the average of the approximation of the variance of the predicted choice probability, which necessitates a second approximation. Third, the fact that $\mathbf{W}_p(\beta)$ depends on β implies that it increases the computational burden in the event a Bayesian optimal design is desired. Finally, any choice probability is always calculated with respect to a certain choice set, involving a certain number of alternatives. Hence, to be able to use the I-optimality criterion based on predicted probabilities, we need to pick a choice set size, a number of choice sets to consider and alternatives to be included in these choice sets. As there is no reason to prefer one choice set size, one number of choice sets, or one set of alternatives over another; there is no indisputable way to define the I-optimality criterion based on choice probabilities. Due to these weaknesses, we prefer to define the I-optimality criterion for choice experiments with mixtures based on predicted utilities.

I-optimality for predicted utilities

In this chapter, we base the I-optimality criterion on predicted utilities. One reason to do so is that that approach does not suffer from the four weaknesses we identified for the I-optimality criterion based on predicted choice probabilities. Another reason for our approach is that, to find the mixture that maximizes the consumer preferences, it suffices to identify the mixture that maximizes the predicted utility. As a matter of fact, the mixture with the largest possible predicted utility will automatically have the largest predicted choice probability in any possible choice set. Finally, starting from predicted utilities is mathematically elegant because there is a closed form expression for the variance of the predicted perceived utility, for a given parameter vector β .

The variance of the predicted utility is defined as

$$\text{Var} \left[\hat{U}_{js} \right] = \mathbf{f}^T(\mathbf{x}_{js}) \text{Var} \left[\hat{\beta} \right] \mathbf{f}(\mathbf{x}_{js}) = \mathbf{f}^T(\mathbf{x}_{js}) \mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{f}(\mathbf{x}_{js}),$$

for a given parameter vector β . As a result, the average variance of the predicted utility is

$$\begin{aligned} \int_{\chi} \mathbf{f}^T(\mathbf{x}_{js}) \mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{f}(\mathbf{x}_{js}) d\mathbf{x}_{js} &= \int_{\chi} \text{tr} \left[\mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{f}(\mathbf{x}_{js}) \mathbf{f}^T(\mathbf{x}_{js}) \right] d\mathbf{x}_{js} \\ &= \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{W}_u] \end{aligned}$$

where

$$\mathbf{W}_u = \int_{\chi} \mathbf{f}(\mathbf{x}_{js}) \mathbf{f}^T(\mathbf{x}_{js}) d\mathbf{x}_{js} \quad (2.8)$$

is the moments matrix corresponding to our definition of the I-optimality criterion, with the subscript u referring to the fact that this is the moments matrix for the I-optimality criterion based on predicted utilities.

In the event the experimental region χ is the $(q - 1)$ -dimensional simplex, there exists a closed-form expression for the moments matrix \mathbf{W}_u that does not depend on the parameter vector β (Goos et al., 2016). This means that, even if a Bayesian optimal design is desired, the moments matrix needs to be computed only once in the design creation process, reducing the computational burden. The elements of the moments matrix \mathbf{W}_u can be obtained using the following formula given in Goos et al. (2016), Goos et al. (2020), and DeGroot (2005):

$$\int_{\chi} x_1^{p_1} x_2^{p_2} \dots x_q^{p_q} dx_1 dx_2 \dots dx_{q-1} = \frac{\prod_{i=1}^q \Gamma(p_i + 1)}{\Gamma(q + \sum_{i=1}^q p_i)}.$$

Our local I-optimality criterion is given by

$$\mathcal{I} = \int_{\chi} \text{Var} [\hat{U}_{js}] d\mathbf{x}_{js} = \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{W}_u],$$

while our Bayesian I-optimality criterion is

$$\mathcal{I}_B = \int_{\mathbb{R}^r} \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{W}_u] \pi(\beta) d\beta. \quad (2.9)$$

In these expressions, we ignore the volume of the experimental region, $1/(q - 1)!$, because, for the optimization of the designs, it is an irrelevant constant.

2.4 Construction of I-optimal designs

To compute Bayesian I-optimal designs for choice experiments with mixtures, we used a coordinate-exchange algorithm (Goos & Jones, 2011; Meyer & Nachtsheim, 1995), modified to generate Bayesian I-optimal designs for choice experiments with mixtures. A coordinate-exchange algorithm was also used by Kessels et al. (2009) and Ruseckaite et al. (2017) in the context of choice experimentation. Our algorithm was implemented in the R programming language (R Core Team, 2017) in which we created a package called `opdesmixr` which is available at <https://github.com/mariobecerra/opdesmixr>. The package was created with the aid of several other R packages (Eddelbuettel, 2013; Eddelbuettel & Balamuta, 2018; Eddelbuettel & François, 2011; Eddelbuettel & Sanderson, 2014; Hamilton & Ferry, 2018; Henry & Wickham, 2020; Wickham, 2016; Wickham et al., 2020), and allows the computation of locally D-optimal, Bayesian D-optimal, locally I-optimal, and Bayesian I-optimal designs for first-order, second-order, and special-cubic Scheffé models. The user must specify either a single parameter vector for a locally optimal design or a matrix of draws from the parameter vector's prior distribution for a Bayesian optimal design. The user must also specify the number of choice sets and the number of alternatives per choice set.

Our coordinate-exchange algorithm starts from a random initial design, and starts by optimizing the first ingredient proportion of the first alternative within the first

choice set, followed by the second ingredient proportion of the first alternative within the first choice set, and so on, until all q ingredient proportions have been optimized. The algorithm then repeats this process for each alternative and each choice set in the design. The whole process is repeated until the design can no longer be improved or until a maximum number of iterations has been reached. As pointed out in Piepel et al. (2005), Goos and Jones (2011), and Ruseckaite et al. (2017), the coordinate-exchange algorithm has to undergo some modifications to deal with mixtures. As a matter of fact, because the mixture proportions have to sum up to one, they cannot be changed independently, and a change in one proportion requires a change in at least one other proportion. We deal with this dependency by changing proportions using the so-called Cox effect direction (Cornell, 2002; Goos & Jones, 2011; Piepel et al., 2005). This means that, after a change of one of the ingredient proportions, x_{ijs} , to $x_{ijs} + \Delta$, we modify the other $q - 1$ proportions as follows:

$$x_{kjs}^{\text{new}} := \begin{cases} \left(1 - \frac{\Delta}{1-x_{ijs}}\right) x_{kjs} & \text{if } x_{ijs} \neq 1, \\ \frac{1-(x_{ijs}+\Delta)}{q-1} & \text{if } x_{ijs} = 1. \end{cases}$$

Three other aspects concerning our coordinate-exchange algorithm are worth mentioning too. First, we recommend running the coordinate-exchange algorithm multiple times, each time starting from a different random initial design. This is because the coordinate-exchange algorithm is a heuristic optimization algorithm, which cannot guarantee optimality, and by running it multiple times, we have a bigger chance of finding a truly optimal design. The larger the number of ingredients and the more complex the model, the larger the number of starts of the coordinate-exchange algorithm should be. For the examples discussed in Section 2.5, we use 80 random starts of the algorithm. The number of ingredients is as low as three and the number of parameters is only six, so this number of initial designs seemed to be enough to avoid local optima since most of the resulting designs had a similar optimality value. For a larger design the number of random starts might have to be increased to avoid this problem. Second, we seek the optimal value of every individual ingredient proportion x_{ijs} using Brent's univariate optimization method (Brent, 1973). Third, we need to approximate the Bayesian optimality criteria numerically, because there is no closed-form solution to the integrals in Equations (2.5) and (2.9). It is common to do this utilizing random or systematic draws from the prior distribution $\pi(\beta)$ (Kessels et al., 2009; Ruseckaite et al., 2017; Train, 2009; Yu et al., 2010). Denoting the R draws from the prior distribution by $\beta^{(i)}$, the approximations for Equations (2.5) and (2.9) are

$$\mathcal{D}_B \approx \frac{1}{R} \sum_{i=1}^R \left[\det \left(\mathbf{I}^{-1}(\mathbf{X}, \beta^{(i)}) \right) \right]^{\frac{1}{r}}, \quad (2.10)$$

and

$$\mathcal{I}_B \approx \frac{1}{R} \sum_{i=1}^R \text{tr} \left[\mathbf{I}^{-1}(\mathbf{X}, \beta^{(i)}) \mathbf{W}_u \right]. \quad (2.11)$$

One commonly used method to obtain draws from a prior distribution involves Halton sequences. The resulting Halton draws provide a good coverage of the

entire density domain, as well as negatively correlated draws that reduce the variance of the approximation to the integral (Train, 2009). Therefore, like Ruseckaite et al. (2017), we used 128 Halton draws from the prior distribution in both of our examples in the next section. The number 128 provides a good enough approximation for the number of parameters in the models used in the two examples. For choice experiments involving more model parameters, a larger number of Halton draws should be used. For more details about Halton draws and other approximation methods, as well as their advantages and disadvantages, we refer to Yu et al. (2010).

2.5 Results

This section shows our computational results for two example choice experiments involving a mixture. The first example is the taste experiment involving cocktails from Courcoux and Séménou (1997b), while the second example involves an experiment with artificial sweeteners for a sports drink from

2.5.1 Cocktail preferences

Ruseckaite et al. (2017) revisited an experiment by Courcoux and Séménou (1997b) in which seven fruit cocktails involving mango juice (whose true proportion we denote by a_1), blackcurrant syrup (whose true proportion we denote by a_2), and lemon juice (whose true proportion we denote by a_3) were tasted. This was done by 60 consumers which were asked to taste different pairs of the seven fruit cocktails and to indicate their preferred cocktail in each pair. Each respondent had to evaluate eight of 21 possible pairs, resulting in a final experimental design with $60 \times 8 = 480$ choice sets of size 2.

In the experiment, Courcoux and Séménou (1997b) imposed lower bounds of 0.3, 0.15 and 0.1 on the three true ingredient proportions a_1 , a_2 , and a_3 . To deal with this issue and to be able to use our implementation of the coordinate-exchange algorithm, like Ruseckaite et al. (2017) did, we expressed the mixtures defining the cocktails in terms of so-called pseudocomponents x_1 , x_2 , and x_3 . These pseudocomponents are defined such that they take a minimum value of 0 and a maximum value of 1, and sum up to one. The conversion of the true ingredient proportions into pseudocomponent proportions is done via the formula $x_i = (a_i - L_i)/(1 - L)$, where L_i denotes the lower bound of ingredient i and L is the sum of the lower bounds for all q ingredient proportions.

To compute D-optimal designs for the cocktail experiment, Ruseckaite et al. (2017) obtained a prior distribution for the parameter vector β in a special-cubic Scheffé model. More specifically, they re-analyzed the data from Courcoux and Séménou (1997b) and derived a multivariate normal prior distribution for β with mean vector

$$\beta_0 = (1.36, 1.57, 2.47, -0.43, 0.50, 1.09)^T$$

and variance-covariance matrix

$$\Sigma_0 = \begin{pmatrix} 6.14 & 5.00 & 2.74 & -0.43 & -2.81 & -3.33 \\ 5.00 & 6.76 & 4.47 & -1.79 & -6.13 & -3.51 \\ 2.74 & 4.47 & 3.45 & -1.38 & -4.71 & -2.17 \\ -0.43 & -1.79 & -1.38 & 1.18 & 2.39 & 0.71 \\ -2.81 & -6.13 & -4.71 & 2.39 & 7.43 & 2.71 \\ -3.33 & -3.51 & -2.17 & 0.71 & 2.71 & 2.49 \end{pmatrix}.$$

We used the same prior distribution to compute optimal designs with 16 choice sets of size 2 assuming a special-cubic Scheffé model. First, we computed a Bayesian D-optimal design to benchmark our implementation of the coordinate-exchange algorithm against that of Ruseckaite et al. (2017), and observed that our design has a slightly better D-optimality criterion value than the original when evaluated using our set of Halton draws. After this validation of our algorithm, we also computed a Bayesian I-optimal design. Our Bayesian D- and I-optimal designs are given in Tables A.9 and A.10 in Appendix A and shown graphically in Figure 2.1. In the figure, the mixtures in each of the 16 choice sets are presented in terms of the pseudocomponent proportions and visualized using different markers for each choice set. The four colored areas in the graph correspond to four prior utility intervals, corresponding to the cutoff values 0, 0.5625, 1.125, 1.6875, and 2.25. The yellow area is the set of mixtures with the lowest a priori utility values, while the red area indicates the set of mixtures with the highest a priori utility values.

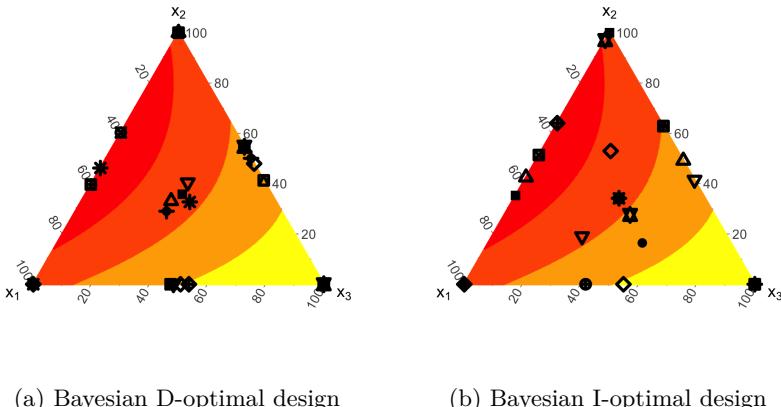


Figure 2.1: Bayesian optimal designs produced by our coordinate-exchange algorithm for the cocktail experiment. The colors represent utilities belonging to the following intervals: ■ [0, 0.5625), ■ [0.5625, 1.125), ■ [1.125, 1.6875), ■ [1.6875, 2.25).

Figure 2.2 shows the fraction of the design space plots of the two Bayesian optimal designs we computed and of the benchmark design from Ruseckaite et al. (2017). Fraction of design space plots were originally introduced by Zahran et al. (2003) and display the performance of a design in terms of the prediction variance for each

point in the experimental region or design space. The horizontal axis corresponds to a fraction of the experimental region, while the vertical axis ranges from the minimum prediction variance to the maximum prediction variance over the entire experimental region (Goos & Jones, 2011). In the context of choice experiments, the prediction variance depends on the unknown parameter vector. We dealt with this issue by computing prediction variances for 128 Halton draws from the prior distribution of the parameter vector and averaging the results.

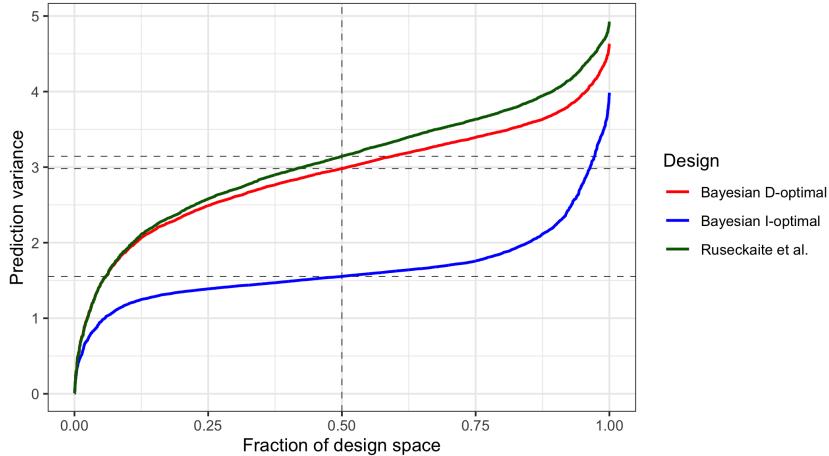


Figure 2.2: Fraction of design space plot of our Bayesian D- and I-optimal designs for the cocktail experiment as well as the Bayesian D-optimal design from Ruseckaite et al. (2017).

The most striking conclusion from the fraction of design space plot is that the prediction variances are much lower for the Bayesian I-optimal design than for the Bayesian D-optimal designs. The plot shows, for instance, that the median prediction variance for the Bayesian I-optimal design is about 1.55, while it is about 3 for our Bayesian D-optimal design and 3.2 for the benchmark design. The maximum prediction variance is also substantially lower for the Bayesian I-optimal design. In summary, using the Bayesian I-optimal design provides much added value in terms of precision of prediction when compared to Bayesian D-optimal designs.

It is not easy to describe the properties of optimal choice designs. To compare the Bayesian D- and I-optimal designs, we quantified the utility balance in the designs' choice sets and computed the Euclidean distances between the alternatives in the choice sets. Utility balance refers to the property that alternatives within a choice set possess the same or a similar a priori utility and therefore have the same or almost the same a priori choice probability. Utility balance was advocated by Huber and Zwerina (1996) as a desirable property for choice designs. A choice set of two alternatives is perfectly utility balanced in the event the choice probabilities of the two alternatives both equal 0.5 and the product of the two probabilities is 0.25. Choice sets that are not at all utility balanced involve alternatives with very different utilities and choice probabilities. For such choice sets, the product

of the two choice probabilities is substantially lower than 0.25. Figure 2.3a shows boxplots of the product of the choice probabilities in the choice sets of our Bayesian D-optimal and our Bayesian I-optimal design. The D-optimal design tends to have choice sets with a higher utility balance than the I-optimal design. However, the median values for the products of the choice probabilities in both designs are very similar and roughly equal to 0.17. This value corresponds to choice probabilities of about 0.78 and 0.22, implying that the Bayesian optimal designs do not exhibit much utility balance. Figure 2.3b shows the Euclidean distances between the two alternatives within a choice set for our Bayesian D- and I-optimal designs. The alternatives within a single choice set tend to be closer together in the D-optimal design than in the I-optimal design. This can also be observed in Figure 2.1 by comparing the distances between the mixtures represented by a given marker. That the alternatives within the choice sets of the I-optimal design are located further from each other is in line with our observations that they exhibit less utility balance. As a matter of fact, when the distance between two mixtures is large, they typically appear in a different utility interval (in other words, in a differently colored area) and their choice probabilities are necessarily quite different. This means that the utility balance is low in choice sets where the alternatives are in different colored areas.

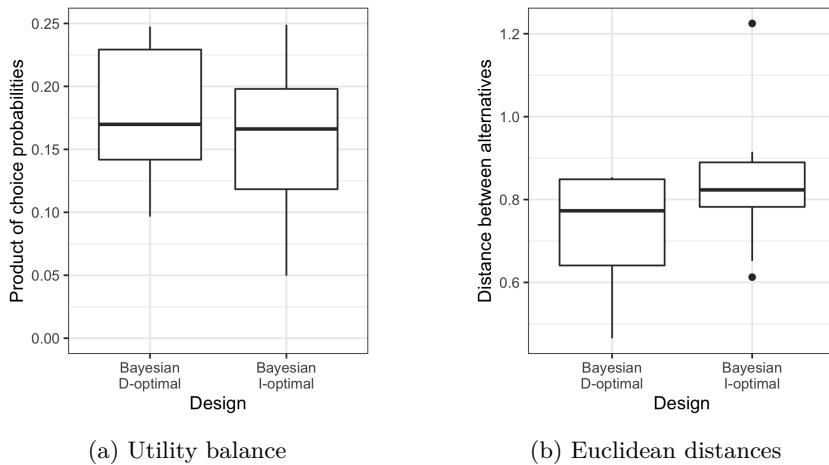


Figure 2.3: Technical properties of our Bayesian D- and I-optimal designs for the cocktail experiment

2.5.2 Artificial sweetener experiment

As a second example, we also revisit the artificial sweetener experiment, a three-ingredient mixture experiment from Cornell (2002), intended to investigate whether an artificial sweetener could be used in an athletic sports drink. The original response variable of interest was ‘intensity of sweetness aftertaste’, but, just like Ruseckaite et al. (2017), we interpret the intensity of sweetness aftertaste as being proportional to the utility in the multinomial logit model and consider a choice

experiment consisting of 7 choice sets of size two as an alternative to the original experiment in Cornell (2002).

For this example, Ruseckaite et al. (2017) started with a special-cubic Scheffé model to construct Bayesian D-optimal designs with a multivariate normal prior distribution with mean vector

$$\boldsymbol{\beta}_0 = (0.86, 0.21, 3.07, 2.34, 3.24, -20.59)^T$$

and variance-covariance matrices of the form $\boldsymbol{\Sigma}_0 = \kappa \mathbf{I}_7$, where κ is a positive scalar that controls the level of uncertainty and \mathbf{I}_7 is the identity matrix of size 7. Using a diagonal variance-covariance matrix facilitates the study of the level of uncertainty through the κ parameter. A higher value of κ indicates a higher level of uncertainty concerning the parameter values. The variance-covariance matrix was transformed to the identified parameter space, as mentioned in Section 2.2.3. The transformed variance-covariance matrix, denoted by $\boldsymbol{\Sigma}'_0$, is

$$\boldsymbol{\Sigma}'_0 = \begin{pmatrix} 2\kappa & \kappa & 0 & 0 & 0 & 0 \\ \kappa & 2\kappa & 0 & 0 & 0 & 0 \\ 0 & 0 & \kappa & 0 & 0 & 0 \\ 0 & 0 & 0 & \kappa & 0 & 0 \\ 0 & 0 & 0 & 0 & \kappa & 0 \\ 0 & 0 & 0 & 0 & 0 & \kappa \end{pmatrix}.$$

With our implementation of the coordinate-exchange algorithm, we first computed Bayesian D-optimal designs for the same κ values as Ruseckaite et al. (2017), namely 0.5, 5, 10 and 30. When comparing our D-optimal designs to those of Ruseckaite et al. (2017) using our sets of 128 Halton draws from the prior distributions, we observed that our designs have D-optimality criterion values very close to those of Ruseckaite et al. (2017), with three designs being slightly better and one being slightly worse. This provides another validation to our algorithm. We also computed Bayesian I-optimal designs for the four κ values. All of our Bayesian D- and I-optimal designs are shown graphically in Figure 2.4. They are given in tabular format in Tables A.1–A.8 in Appendix A. The four different colors in Figure 2.4 correspond to four intervals for the utility of the mixtures, with bounds 0, 0.375, 0.75, 1.125, and 1.34. It can be seen that the spread in the points in the optimal designs increases with κ , but this phenomenon is more pronounced for the Bayesian I-optimal designs than for the Bayesian D-optimal designs.

Figure 2.5 shows the fraction of design space plots for our Bayesian D- and I-optimal designs, as well as for the designs of Ruseckaite et al. (2017). For each value of κ , the I-optimal design has a much lower prediction variance than our D-optimal design and that of Ruseckaite et al. (2017). The difference in predictive performance, in favor of the Bayesian I-optimal designs, increases with κ .

Figure 2.6 shows boxplots of the product of the choice probabilities in each choice set for our Bayesian D- and I-optimal designs for the different κ values. For $\kappa = 0.5$, both types of designs score highly in utility balance, and there is hardly any difference between them in terms of utility balance. However, as κ increases, the product of the probabilities drops, meaning that the designs become less utility balanced.

Figure 2.7 shows boxplots of the Euclidean distances between the two alternatives within a choice set for our Bayesian D- and I-optimal designs. Just like in the cocktail experiment, mixtures within a single choice set tend to be closer together in the D-optimal designs than in the I-optimal designs, except when $\kappa = 30$. For that value of κ , there is no major difference in the Euclidean distances between the alternatives within the choice sets of the D- and I-optimal designs. Finally, Figure 2.7 shows that the distances between alternatives with choice sets decrease with the value of κ .

2.6 Discussion

In this chapter, we introduced a computationally efficient definition for I-optimal designs for choice experiments and embedded the new I-optimality criterion in a coordinate-exchange algorithm for constructing I-optimal designs. By means of two examples from the literature, we demonstrated that the I-optimal designs perform substantially better than their D-optimal counterparts in terms of the variance of the predicted utility. We observed that I-optimal designs do not possess the utility balance property, which Huber and Zwerina (1996) considered to be desirable for efficient choice designs. However, Louviere et al. (2011) argued against it, suggesting to instead derive an optimal design using an appropriate prior distribution. We believe this to be a more sensible choice given that it is our interest to have designs that yield precise predictions for any combination of ingredient proportions.

An immediate possible extension of this work comes to the light when we consider that the preference for a mixture may depend on characteristics other than its composition. For example, the ideal cocktail composition may also depend on the temperature at which it is served, or the most preferred bread might not only depend on the proportions of the various ingredients, but also on the baking time and the baking temperature. One practical example of such a scenario can be found in Zijlstra et al. (2019), who observed that the preferred mobility budget mixture depends on the budget size. To cope with this kind of complication, our choice model for mixtures must be extended to deal with the additional characteristics, typically called *process variables* (Goos & Jones, 2011). This extension is shown in Chapter 3.

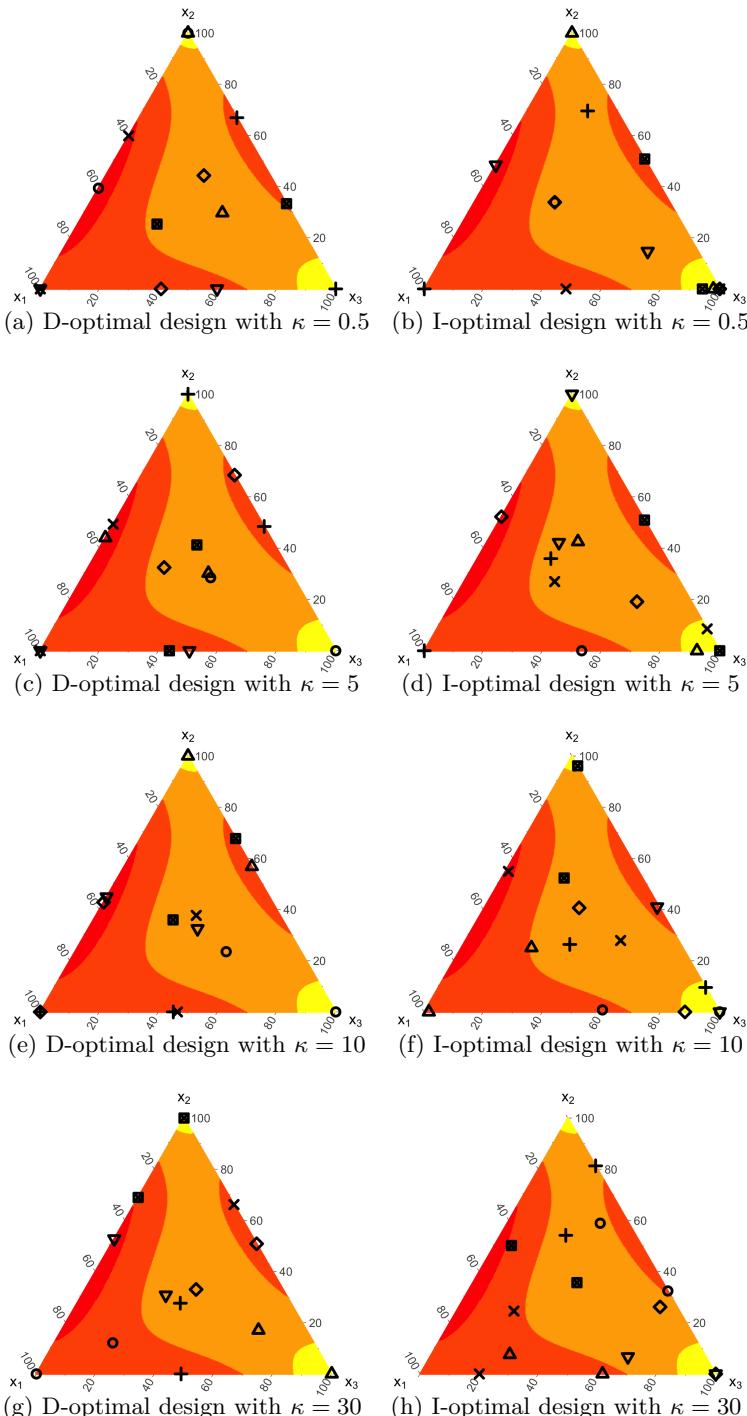


Figure 2.4: Bayesian optimal designs for the artificial sweetener experiment. The colors represent utilities belonging to the following intervals: [0, 0.375), [0.375, 0.75), [0.75, 1.125), [1.125, 1.34).

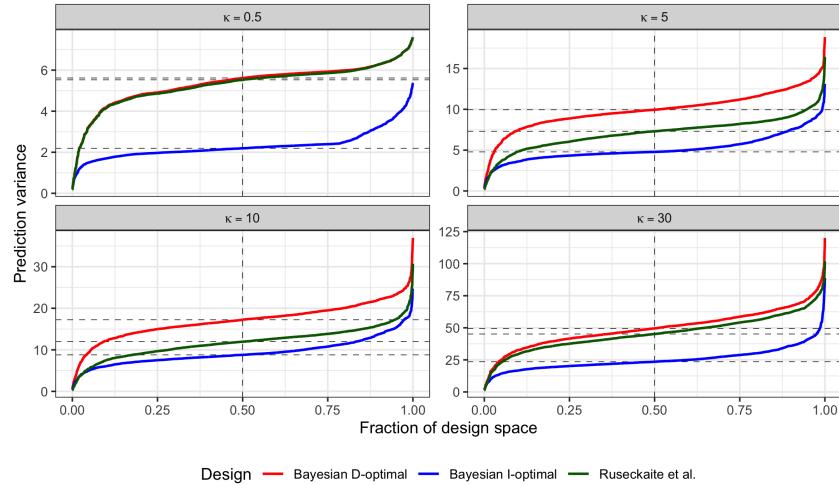


Figure 2.5: Fraction of design space plots of the Bayesian optimal designs for the artificial sweetener experiment

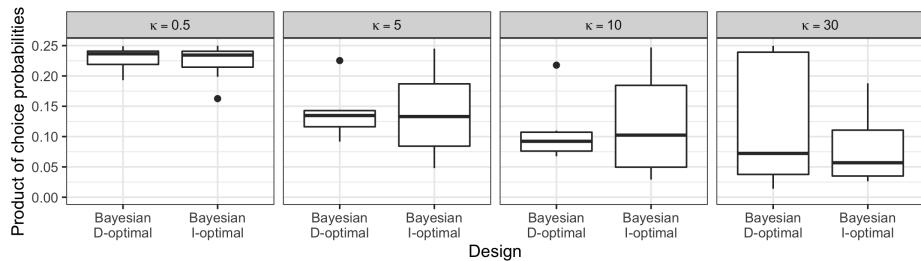


Figure 2.6: Measures of utility balance for our Bayesian D- and I-optimal designs for the artificial sweetener experiment

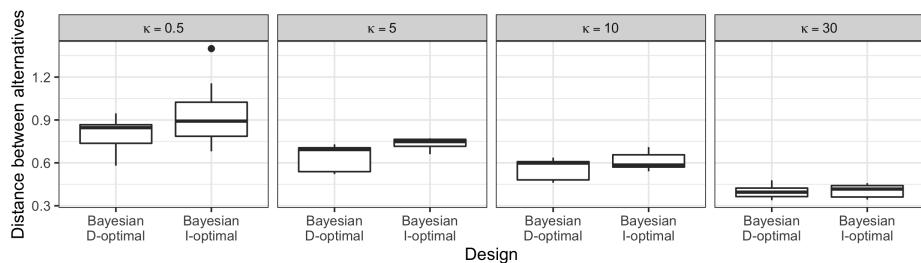


Figure 2.7: Euclidean distances between alternatives within a choice set for our Bayesian D- and I-optimal designs for the artificial sweetener experiment

Chapter 3

Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables¹

Abstract

Many food products involve mixtures of ingredients, where the mixtures can be expressed as combinations of ingredient proportions. In many cases, the quality and the consumer preference may also depend on the way in which the mixtures are processed. The processing is generally defined by the settings of one or more process variables. Experimental designs studying the joint impact of the mixture ingredient proportions and the settings of the process variables are called mixture-process variable experiments. In this article, we show how to combine mixture-process variable experiments and discrete choice experiments, to quantify and model consumer preferences for food products that can be viewed as processed mixtures. First, we describe the modeling of data from such combined experiments. Next, we describe how to generate D- and I-optimal designs for choice experiments involving mixtures and process variables, and we compare the two kinds of designs using two examples.

¹This chapter is, with minor adaptations, a copy of the following article: Becerra, Mario & Goos, Peter (2023) *Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables*, Food Quality and Preference, 110, 104928, DOI: 10.1016/j.foodqual.2023.104928

3.1 Introduction

As pointed out in the review paper on the state of the art of discrete choice experiments in food research by Lizin et al. (2022), there has been a steady increase in the number of publications on the use of discrete choice experiments concerning food since 2000. A large number of discrete choice experiments in these papers deal with food safety or safety risks, origin or traceability, health or nutrition, biotechnology or genetic modification and animal welfare. The product categories mainly involved meat (beef, pork, poultry, and processed meat products), organic foods, functional foods and foods with nutrition or health claims. In recent years, alternatives to conventional meat received increasing attention. Lizin et al. (2022) also mention that a limited number of choice experiments in published papers were concerned with wine, olive oil, eggs and vegetables.

Despite the fact that many food products involve mixtures of ingredients, publications concerning food-related choice experiments with mixtures are scarce. The first known application of a discrete choice experiment concerning mixtures was published by Courcoux and Séménou (1997b), who modeled the preferences for cocktails involving different proportions of mango juice, lime juice, and blackcurrant syrup. Goos and Hamidouche (2019) defined a way to combine Scheffé models for data from mixture experiments with the logit type models typically used for choice experiments, and presented an alternative analysis of the data from Courcoux and Séménou (1997b). Ruseckaite et al. (2017) and us in Chapter 2 demonstrated how D- and I-optimal designs can be generated for choice experiments with mixtures, applied their work to the cocktail experiment and used an additional example concerning a sports drink.

As witnessed by many of the examples in Cornell (2002), the quality of food products involving mixtures of ingredients often also depends on characteristics unrelated to the composition of the mixture. For example, the firmness of a fish patty depends not only on the types of fish used, but also on baking temperature, baking time, and frying time. The color, aroma, taste, texture and mouthfeel of *pastillas de leche*, a popular Filipino candy, depend on baking time and temperature in addition to mixture ingredients such as cornstarch, flour, glucose, sugar and milk (Apellado-Buenaventura & Valmorida, 2021). The aroma, hardness, crispness, color and fracture force of apple biscuits are affected by the mixture ingredients and the microwave blanching of the apples (Skaltsi et al., 2022). In the general literature on mixture experiments, variables such as baking temperature, baking time, frying time, serving temperature, and microwave blanching are typically called *process variables* (Goos & Jones, 2011).

The fact that the quality of food products involving mixtures depends on the settings of such process variables implies that consumer preferences for these kinds of products will also be impacted by the process variables' settings. For this reason, in this article, we develop the methodology required to perform discrete choice experiments involving mixtures as well as process variables. First, we present a parsimonious model for data from choice experiments with mixture and process variables. Next, we discuss how to generate D- and I-optimal designs for such choice experiments. We discuss D-optimal designs because the D-optimality criterion is

the most popular criterion for designing choice experiments; and I-optimal designs because they focus on precise predictions and precise predictions are helpful to find the optimal mixture formulation in combination with optimal settings for the process variables.

The rest of the chapter is organized as follows. In Section 3.2, we introduce the most often used models for mixture experiments with process variables, the multinomial logit model for choice data and the combination of these two to model choice data concerning mixtures and process variables. In Section 3.3, we discuss the two most commonly used metrics to measure the quality of experimental designs. In Section 3.4, we present some of our computational results and provide example designs for two choice experiments involving a mixture and one or more process variables. Finally, in Section 3.5, we summarize our work.

3.2 Models

In this section, we introduce the most commonly used models for data from mixture experiments with process variables as well as the multinomial logit model for choice data, and explain how to combine the two models for data from choice experiments involving mixtures and process variables.

3.2.1 Models for data from mixture experiments including process variables

Mixture experiments involve two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture. Each mixture is described as a combination of q ingredient proportions, with the constraint that these proportions sum up to one. Due to this constraint, a classical regression model involving an intercept and linear terms in the ingredient proportions exhibits perfect collinearity. Therefore, researchers must use dedicated regression models when analyzing data from mixture experiments. The most commonly used family of models for data from mixture experiments is the Scheffé family (Scheffé, 1958; Scheffé, 1963). The most popular Scheffé models are the first-order, second-order, and special-cubic models.

Denoting the response in a traditional mixture experiment with a continuous outcome by Y and the q ingredient proportions by x_1, x_2, \dots, x_q , with $x_i \geq 0$ and $\sum_{i=1}^q x_i = 1$, the first-order Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \varepsilon. \quad (3.1)$$

The second-order Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \varepsilon, \quad (3.2)$$

and, finally, the special-cubic Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_i x_k x_l + \varepsilon. \quad (3.3)$$

In all three cases, ε denotes the error term, which, for continuous outcomes, is typically assumed to be normally distributed.

In certain experiments involving mixtures, additional factors that might affect the response are studied as well. Generally, these factors describe how the mixture is processed (where the word ‘processed’ should be interpreted in a broad sense). These additional factors are therefore referred to as process variables, and the resulting experiments are called mixture-process variable experiments. For instance, a dough needs to be baked at a certain temperature for a certain time, while the cocktails from the example in Section 3.4.1 need to be cooled to a certain temperature before being served, and the fish patties from the example in Section 3.4.2 are cooked and fried for a specific time at a specific temperature.

Models that involve q mixture ingredients and r process variables can be obtained by combining Scheffé models for the ingredient proportions with response surface models for the process variables (Cornell and Gorman, 1984; Cornell, 1988; Kowalski et al., 2000; Goos and Jones, 2011). For example, consider the second-order Scheffé model in Equation (3.2) for q ingredients x_1, x_2, \dots, x_q and a main-effects-plus-two-factor-interaction model for r process variables z_1, z_2, \dots, z_r defined as

$$Y = \alpha_0 + \sum_{k=1}^r \alpha_k z_k + \sum_{k=1}^{r-1} \sum_{l=k+1}^r \alpha_{kl} z_k z_l + \varepsilon. \quad (3.4)$$

One combined model crosses the terms in Equation (3.2) with each of those in Equation (3.4):

$$\begin{aligned} Y = & \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \sum_{i=1}^q \sum_{l=1}^r \gamma_{il} x_i z_l + \sum_{i=1}^q \sum_{l=1}^{r-1} \sum_{m=l+1}^r \gamma_{ilm} x_i z_l z_m \\ & + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \sum_{l=1}^r \delta_{ikl} x_i x_k z_l + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \sum_{l=1}^{r-1} \sum_{m=l+1}^r \delta_{iklm} x_i x_k z_l z_m + \varepsilon. \end{aligned} \quad (3.5)$$

This model allows the effects of both the ingredient proportions and process variables to jointly affect the response variable. In other words, the model allows the effects of the process variables to depend on the ingredient proportions and the effects of the ingredient proportions to depend on the process variables. The combined model in Equation (3.5) does not include any main effects of the process variables z_1, \dots, z_r . This is because their inclusion would result in an inestimable model due to perfect collinearity. In the event that the effects of the process variables do not depend on the ingredient proportions, all γ_{ik} as well as all γ_{ikl} in the combined model are equal and all δ_{ijk} and all δ_{ijkl} are zero. The model then

simplifies to

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \sum_{k=1}^r \alpha_k z_k + \sum_{k=1}^{r-1} \sum_{l=k+1}^r \alpha_{kl} z_k z_l + \varepsilon. \quad (3.6)$$

This alternative model also combines the models in Equations (3.2) and (3.4), but without crossing any of the terms. Depending on the application, it may be necessary to extend the above models by including cubic terms involving the mixture ingredient proportions (as in the special-cubic Scheffé model in Equation (3.3)) or quadratic terms in the process variables. An example of such an extended model would be

$$\begin{aligned} Y = & \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \beta_{ik} x_i x_k + \sum_{i=1}^{q-2} \sum_{k=i+1}^{q-1} \sum_{l=k+1}^q \beta_{ikl} x_i x_k x_l \\ & + \sum_{i=1}^q \sum_{l=1}^r \gamma_{il} x_i z_l + \sum_{i=1}^q \sum_{l=1}^{r-1} \sum_{m=l+1}^r \gamma_{ilm} x_i z_l z_m \\ & + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \sum_{l=1}^r \delta_{ikl} x_i x_k z_l + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \sum_{l=1}^{r-1} \sum_{m=l+1}^r \delta_{iklm} x_i x_k z_l z_m + \sum_{i=1}^r \alpha_i z_i^2 + \varepsilon. \end{aligned} \quad (3.7)$$

A problem with the combined model in Equation (3.5) is that its number of parameters quickly increases with the number of mixture ingredients and process variables: for q mixture ingredients and r process variables, the total number of parameters is $[q + q(q - 1)/2] \times [1 + r + r(r - 1)/2]$. The extended model in Equation (3.7) even involves $q(q - 1)(q - 2)/6 + r$ extra parameters. In contrast, the model described in Equation (3.6) involves a number of parameters that is as low as $[q + q(q - 1)/2] + [r + r(r - 1)/2]$. The drawback of the latter model is that it may not be realistic. For this reason, Kowalski et al. (2000) suggest a compromise model involving $q + q(q - 1)/2 + qr + r(r - 1)/2 + r$ terms:

$$Y = \sum_{i=1}^q \gamma_i^0 x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_i x_k + \sum_{i=1}^q \sum_{k=1}^r \gamma_i^k x_i z_k + \sum_{i=1}^{r-1} \sum_{k=i+1}^r \alpha_{ik} z_i z_k + \sum_{i=1}^r \alpha_i z_i^2 + \varepsilon. \quad (3.8)$$

Because this compromise model strikes a balance between the overly complex models in Equations (3.5) and (3.7) and the overly simple model in Equation (3.6), we use it as our starting point for computing optimal designs for choice experiments involving mixtures and process variables in the remainder of this chapter.

3.2.2 Multinomial logit model for choice data

The multinomial logit model builds on random-utility theory and assumes that a respondent in a choice experiment faces S choice sets involving J alternatives. The model assumes that, within each choice set $s \in \{1, \dots, S\}$, each respondent chooses the alternative that has the highest perceived utility. Therefore, the probability

that a respondent chooses alternative $j \in \{1, \dots, J\}$ in choice set s , denoted by p_{js} , is the probability that the perceived utility of alternative j in choice set s , denoted by U_{js} , is larger than that of the other alternatives in the choice set:

$$p_{js} = \mathbb{P}[U_{js} > \max(U_{1s}, \dots, U_{j-1,s}, U_{j+1,s}, \dots, U_{Js})].$$

Since, generally, each alternative in a choice set has a set of observable attributes that characterize it, the perceived utility U_{js} can be expressed as

$$U_{js} = \mathbf{f}^T(\mathbf{a}_{js})\boldsymbol{\theta} + \varepsilon_{js}, \quad (3.9)$$

where \mathbf{a}_{js} is the vector that contains the attributes corresponding to alternative j in choice set s , $\mathbf{f}(\mathbf{a}_{js})$ represents the model expansion of this attribute vector, and $\boldsymbol{\theta}$ is the vector containing the model parameters. The model parameters contained within $\boldsymbol{\theta}$ express the preferences of the respondents for the alternatives' attributes. In the multinomial logit model, the error terms ε_{js} are assumed to be independent and identically Gumbel distributed. The Gumbel distribution is also known as the generalized extreme value distribution of type I and as the log-Weibull distribution. As a result of the distributional assumption, it can be shown that

$$p_{js} = \frac{\exp[\mathbf{f}^T(\mathbf{a}_{js})\boldsymbol{\theta}]}{\sum_{t=1}^J \exp[\mathbf{f}^T(\mathbf{a}_{ts})\boldsymbol{\theta}]} \quad (3.10)$$

3.2.3 Model for choice data concerning mixtures and process variables

In this chapter, we focus on choice experiments involving mixtures and process variables. Therefore, we assume that the attributes of the alternatives in the experiments are the proportions of the ingredients of a mixture and the settings of the process variables. Consequently, we assume that the attribute vector \mathbf{a}_{js} from Equation (3.9) contains the q ingredient proportions x_1, x_2, \dots, x_q and the r process variables z_1, \dots, z_r of the j -th alternative in choice set s and that $\mathbf{f}(\mathbf{a}_{js})$ represents the model expansion of these proportions and process variables. As a proof of concept, in this chapter we base the polynomial expansion $\mathbf{f}(\mathbf{a}_{js})$ on a model combining a second-order Scheffé model for the q ingredients in the mixture with a main-effects-plus-two-factor-interaction model for the r process variables, as in Equation (3.8).

When starting from the main-effects-plus-two-factor-interaction model in Equation (3.8), the most natural thing to do would be to write the perceived utility U_{js} of alternative j in choice set s as

$$\begin{aligned} U_{js} = & \sum_{i=1}^q \gamma_i^0 x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} + \sum_{i=1}^r \sum_{k=1}^q \gamma_k^i x_{kjs} z_{ijs} \\ & + \sum_{i=1}^{r-1} \sum_{k=i+1}^r \alpha_{ik} z_{ijs} z_{kjs} + \sum_{i=1}^r \alpha_i z_{ijs}^2 + \varepsilon_{js}, \end{aligned}$$

where x_{ijs} denotes the proportion of the i -th mixture ingredient in alternative j from choice set s , and z_{kjs} denotes the setting of the k -th process variable for alternative j in choice set s , and the error terms ε_{js} are assumed to be independent and identically Gumbel distributed. However, as explained by Ruseckaite et al. (2017), Goos and Hamidouche (2019), and in Chapter 2, due to the constraint that the ingredient proportions sum up to one, this leads to an inestimable multinomial logit model. As a consequence of the constraint, we can rewrite x_{qjs} as $1 - x_{1js} - \dots - x_{q-1,js}$ and U_{js} as

$$\begin{aligned} U_{js} &= \sum_{i=1}^{q-1} \gamma_i^0 x_{ijs} + \gamma_q^0 (1 - x_{1js} - \dots - x_{q-1,js}) + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^r \sum_{k=1}^q \gamma_k^i x_{kjs} z_{ijs} + \sum_{i=1}^{r-1} \sum_{k=i+1}^r \alpha_{ik} z_{ijs} z_{kjs} + \sum_{i=1}^r \alpha_i z_{ijs}^2 + \varepsilon_{js} \\ &= \gamma_q^0 + \sum_{i=1}^{q-1} (\gamma_i^0 - \gamma_q^0) x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^r \sum_{k=1}^q \gamma_k^i x_{kjs} z_{ijs} + \sum_{i=1}^{r-1} \sum_{k=i+1}^r \alpha_{ik} z_{ijs} z_{kjs} + \sum_{i=1}^r \alpha_i z_{ijs}^2 + \varepsilon_{js}. \end{aligned}$$

This final expression for the perceived utility U_{js} involves a constant, γ_q^0 . Since the multinomial logit model only takes into account differences in utility, that constant causes the model to be ill-defined and, hence, inestimable. This can be circumvented by dropping γ_q^0 , defining the parameters $\gamma_i^{0*} = \gamma_i^0 - \gamma_q^0$ for $i \in \{1, \dots, q-1\}$, and using the following expression for the perceived utility:

$$\begin{aligned} U_{js} &= \sum_{i=1}^{q-1} \gamma_i^{0*} x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^r \sum_{k=1}^q \gamma_k^i x_{kjs} z_{ijs} + \sum_{i=1}^{r-1} \sum_{k=i+1}^r \alpha_{ik} z_{ijs} z_{kjs} + \sum_{i=1}^r \alpha_i z_{ijs}^2 + \varepsilon_{js}. \end{aligned} \tag{3.11}$$

The parameter vector $\boldsymbol{\theta}$ then becomes

$$\begin{aligned} \boldsymbol{\theta} &= (\gamma_1^{0*}, \gamma_2^{0*}, \dots, \gamma_{q-1}^{0*}, \gamma_{1,2}^0, \dots, \gamma_{q-1,q}^0, \gamma_1^1, \gamma_2^1, \dots, \gamma_q^1, \gamma_{q,1}^2, \dots, \gamma_q^r, \\ &\quad \alpha_{1,2}, \dots, \alpha_{r-1,r}, \alpha_1, \dots, \alpha_r)^T. \end{aligned}$$

This vector has $q + \frac{q(q-1)}{2} + qr + \frac{r(r-1)}{2} + r - 1$ elements.

3.3 Optimal design criteria

In the literature on the optimal design of choice experiments in general, several criteria have been studied. Kessels et al. (2006) elaborate on the D-, I-, A-, and G-

optimality criteria for the multinomial logit model and compare the performances of the resulting choice designs. However, in the literature on optimal design of choice experiments with mixtures, the two optimality metrics that have been studied are D-optimality and I-optimality. In this section, we extend the D- and I-optimality criteria to cope with the multinomial logit model for choice experiments involving mixtures as well as process variables.

3.3.1 Information matrix

In order to create D- and I-optimal experimental designs, we need to compute a design's information matrix corresponding to the model under investigation. For the multinomial logit model, the information matrix depends on the unknown parameter vector $\boldsymbol{\theta}$ through the choice probabilities p_{js} defined in Equation (3.10). This is typical for models that are not linear in the parameters, such as discrete choice models, and it implies that prior information is needed to find optimal designs. This information can be provided in the form of a point estimate, or in the form of a prior distribution (Atkinson and Haines, 1996; Kessels et al., 2006; Ruseckaite et al., 2017). The use of a point estimate leads to so-called locally optimal designs, which have the problem that they may perform poorly for values of the parameter vector $\boldsymbol{\theta}$ for which they were not optimized. This weakness of locally optimal designs is, of course, highly relevant given that the true values of the model parameters are unknown. An alternative is to use a prior distribution, which leads to so-called Bayesian optimal designs. In addition to taking into account prior information, Bayesian optimal designs also take into account the uncertainty about the parameter vector $\boldsymbol{\theta}$ through the use of a prior distribution $\pi(\boldsymbol{\theta})$ that summarizes the prior knowledge concerning the parameter vector $\boldsymbol{\theta}$.

The information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\theta})$ for the multinomial logit model is the sum of the information matrices of each of the S choice sets (Kessels et al., 2006):

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\theta}) = \sum_{s=1}^S \mathbf{X}_s^T (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T) \mathbf{X}_s,$$

with $\mathbf{p}_s = (p_{1s}, \dots, p_{Js})^T$, $\mathbf{P}_s = \text{diag}(\mathbf{p}_s)$, $\mathbf{X}_s^T = [\mathbf{f}(\mathbf{a}_{1s}), \mathbf{f}(\mathbf{a}_{2s}), \dots, \mathbf{f}(\mathbf{a}_{Js})]$ the model matrix containing the model expansions of the attribute levels of all J alternatives in choice set s , and $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_S]$ the model matrix for all S choice sets. The inverse of the information matrix is the asymptotic variance-covariance matrix of the maximum likelihood estimates of the parameter vector $\boldsymbol{\theta}$.

3.3.2 D-optimal designs

For a model matrix \mathbf{X} and prior parameter vector $\boldsymbol{\theta}$, the D-optimality criterion can be defined as

$$\mathcal{D} = [\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta}))]^{\frac{1}{m}}, \quad (3.12)$$

where $\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta})$ is the inverse of the information matrix and m is the number of parameters in the model. A D-optimal design minimizes the \mathcal{D} -value. Since

the D-optimal design approach focuses on minimizing the generalized variance of the maximum likelihood estimators of the model parameters, it can be viewed as an estimation-based approach. D-optimality is arguably the most traditional metric used in the literature on the design of choice experiments (Bliemer et al., 2009; Bliemer and Rose, 2010; Bliemer and Rose, 2011; Burgess and Street, 2005; Grasshoff et al., 2003; Kessels et al., 2011).

The definition in Equation (3.12) uses a prior point estimate of the parameter vector $\boldsymbol{\theta}$. However, as we mentioned above, a prior distribution can also be used to obtain a Bayesian D-optimal design. The Bayesian D-optimality criterion is generally defined in the literature as the average of the D-optimality criterion over the prior distribution (Bliemer et al., 2009; Bliemer and Rose, 2011; Kessels et al., 2011). Therefore, following Ruseckaite et al. (2017) and Chapter 2, we define the Bayesian D-optimality criterion for the multinomial logit model as

$$\mathcal{D}_B = \int_{\mathbb{R}^m} [\det(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta}))]^{\frac{1}{m}} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (3.13)$$

where $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$. Note that we call a design that minimizes the expression in Equation (3.13) a Bayesian D-optimal design, even though the criterion does not take into account the posterior distribution and some authors therefore prefer to call such a design a pseudo-Bayesian design (e.g., Ryan et al. (2016)).

3.3.3 I-optimal designs

The I-optimality criterion is generally defined as the average prediction variance over the experimental region, which is why it can be seen as a prediction-oriented criterion: it focuses on getting precise predictions with the estimated statistical model. I-optimality is also sometimes called V-optimality (Goos and Syafitri, 2014; Kessels et al., 2006).

When using choice models, there are two ways in which we can define I-optimality. If the goal is to predict choice probabilities, the I-optimality criterion is the average variance of the predicted choice probabilities. If the goal is to predict perceived utilities, the I-optimality criterion is the average variance of the predicted utilities. In Chapter 2, we introduced a computationally efficient definition for I-optimal designs for choice experiments focused on the perceived utilities. This is the definition we will use here too. Under this definition, the I-optimality criterion is

$$\mathcal{I} = \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta}) \mathbf{W}], \quad (3.14)$$

where $\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta})$ again denotes the inverse of the information matrix for model matrix \mathbf{X} and prior parameter vector $\boldsymbol{\theta}$. The matrix \mathbf{W} is the moments matrix, defined as

$$\mathbf{W} = \int_{\chi} \mathbf{f}(\mathbf{a}_{js}) \mathbf{f}^T(\mathbf{a}_{js}) d\mathbf{a}_{js}, \quad (3.15)$$

with $\mathbf{f}(\mathbf{a}_{js})$ again the model expansion of attribute vector \mathbf{a}_{js} and χ the experimental region which combines the $(q - 1)$ -dimensional simplex S_{q-1} for

the ingredient proportions and an r -dimensional hyperrectangle for the possible settings of the process variables.

To compute the moments matrix \mathbf{W} for the model described in Equation (3.11), we first need to compute the matrix $\mathbf{f}(\mathbf{a}_{js})\mathbf{f}^T(\mathbf{a}_{js})$, which has elements of the form

$$\left(\prod_{k=1}^q x_k^{n_k} \right) \left(\prod_{l=1}^r z_l^{m_l} \right),$$

for some $n_k, m_l \in \mathbb{N}$, $k \in \{1, \dots, q\}$ and $l \in \{1, \dots, r\}$. Hence, each element of the moments matrix is of the form

$$\int_{\chi} \left(\prod_{k=1}^q x_k^{n_k} \right) \left(\prod_{l=1}^r z_l^{m_l} \right) dx_1 \dots dx_q dz_1 \dots dz_r,$$

which can be separated in two parts: one corresponding to the process variables and one part corresponding to the ingredient proportions. Therefore, assuming that the r process variables take values from the intervals $[a_1, b_1], [a_2, b_2], \dots, [a_r, b_r]$, the i -th element in the j -th column of the moments matrix, denoted by W_{ij} , can be calculated as

$$\begin{aligned} W_{ij} &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_r}^{b_r} \prod_{l=1}^r z_l^{m_l} \left(\int_{S_{q-1}} \prod_{k=1}^q x_k^{n_k} dx_1 \dots dx_{q-1} \right) dz_1 \dots dz_r, \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_r}^{b_r} \prod_{l=1}^r z_l^{m_l} \left(\frac{\prod_{k=1}^q n_k!}{(q-1 + \sum_{k=1}^q n_k)!} \right) dz_1 \dots dz_r, \\ &= \left(\prod_{l=1}^r \frac{b_l^{m_l+1} - a_l^{m_l+1}}{m_l + 1} \right) \left(\frac{\prod_{k=1}^q n_k!}{(q-1 + \sum_{k=1}^q n_k)!} \right). \end{aligned}$$

If we adopt the convention that the settings of the process variables are rescaled to the $[-1, +1]$ interval, the hyperrectangle becomes a hypercube and the expression for W_{ij} can be simplified to

$$W_{ij} = \left(\prod_{l=1}^r \frac{1^{m_l+1} - (-1)^{m_l+1}}{m_l + 1} \right) \left(\frac{\prod_{k=1}^q n_k!}{(q-1 + \sum_{k=1}^q n_k)!} \right). \quad (3.16)$$

In the event one of the m_l values is odd, $1^{m_l+1} - (-1)^{m_l+1}$ is zero and W_{ij} also becomes zero. In the event all m_l values are even, $1^{m_l+1} - (-1)^{m_l+1}$ is equal to 2.

So, for example, in the case where there are three mixture variables and one process variable (i.e., $q = 3$ and $r = 1$) the model expansion $\mathbf{f}(\mathbf{a}_{js})$ is $(x_1, x_2, x_1x_2, x_1x_3, x_2x_3, x_1z, x_2z, x_3z, z^2)^T$. Multiplying $\mathbf{f}(\mathbf{a}_{js})$ by its transpose

yields the matrix $\mathbf{f}(\mathbf{a}_{js})\mathbf{f}^T(\mathbf{a}_{js})$

$$\begin{bmatrix} x_1^2 & x_1x_2 & x_1^2x_2 & x_1^2x_3 & x_1x_2x_3 & x_1^2z & x_1x_2z & x_1x_3z & x_1z^2 \\ x_1x_2 & x_2^2 & x_1x_2^2 & x_1x_2x_3 & x_2^2x_3 & x_1x_2z & x_2^2z & x_2x_3z & x_2z^2 \\ x_1^2x_2 & x_1x_2^2 & x_1^2x_2^2 & x_1^2x_2x_3 & x_1x_2^2x_3 & x_1^2x_2z & x_1x_2^2z & x_1x_2x_3z & x_1x_2z^2 \\ x_1^2x_3 & x_1x_2x_3 & x_1^2x_2x_3 & x_1^2x_3^2 & x_1x_2x_3^2 & x_1^2x_3z & x_1x_2x_3z & x_1x_3^2z & x_1x_3z^2 \\ x_1x_2x_3 & x_2^2x_3 & x_1x_2^2x_3 & x_1x_2x_3^2 & x_2^2x_3 & x_1x_2x_3z & x_2^2x_3z & x_2x_3^2z & x_2x_3z^2 \\ x_1^2z & x_1x_2z & x_1^2x_2z & x_1^2x_3z & x_1x_2x_3z & x_1^2z^2 & x_1x_2z^2 & x_1x_3z^2 & x_1z^3 \\ x_1x_2z & x_2^2z & x_1x_2^2z & x_1x_2x_3z & x_2^2x_3z & x_1x_2z^2 & x_2^2z^2 & x_2x_3z^2 & x_2z^3 \\ x_1x_3z & x_2x_3z & x_1x_2x_3z & x_1x_3^2z & x_2x_3^2z & x_1x_3z^2 & x_2x_3z^2 & x_3^2z^2 & x_3z^3 \\ x_1z^2 & x_2z^2 & x_1x_2z^2 & x_1x_3z^2 & x_2x_3z^2 & x_1z^3 & x_2z^3 & x_3z^3 & z^4 \end{bmatrix}.$$

To illustrate how W_{11} is calculated, we start from the first element in the first row and the first column in this matrix, i.e., x_1^2 . This term is the square of the first mixture ingredient proportion. Hence, its exponent n_1 is equal to 2. The other two mixture variables, x_2 and x_3 , are not present, meaning their exponents n_2 and n_3 are 0. Additionally, this element does not involve any process variables, meaning $m_1 = 0$. Using Equation (3.16), we obtain

$$\begin{aligned} W_{11} &= \left(\frac{1^{m_1+1} - (-1)^{m_1+1}}{m_1 + 1} \right) \left(\frac{n_1! \times n_2! \times n_3!}{(3 - 1 + n_1 + n_2 + n_3)!} \right) \\ &= \left(\frac{1^{0+1} - (-1)^{0+1}}{0 + 1} \right) \left(\frac{2! \times 0! \times 0!}{(3 - 1 + 2 + 0 + 0)!} \right) = \left(\frac{2}{1} \right) \left(\frac{2}{24} \right) = \frac{1}{6}. \end{aligned}$$

As another illustration, we calculate W_{99} . To this end, we start from the element in the last row and the last column of $\mathbf{f}(\mathbf{a}_{js})\mathbf{f}^T(\mathbf{a}_{js})$, i.e., z^4 . This term is the process variable raised to the 4-th power. Hence, $m_1 = 4$. None of the mixture variables are present, meaning that their exponents are all 0, and thus $n_1 = n_2 = n_3 = 0$. So, using Equation (3.16) again, we obtain

$$\begin{aligned} W_{99} &= \left(\frac{1^{m_1+1} - (-1)^{m_1+1}}{m_1 + 1} \right) \left(\frac{n_1! \times n_2! \times n_3!}{(3 - 1 + n_1 + n_2 + n_3)!} \right) \\ &= \left(\frac{1^{4+1} - (-1)^{4+1}}{4 + 1} \right) \left(\frac{0! \times 0! \times 0!}{(3 - 1 + 0 + 0 + 0)!} \right) = \left(\frac{2}{5} \right) \left(\frac{1}{24} \right) = \frac{1}{5}. \end{aligned}$$

Following this process for each of the elements in the matrix $\mathbf{f}(\mathbf{a}_{js})\mathbf{f}^T(\mathbf{a}_{js})$, we obtain the full moments matrix,

$$\mathbf{W} = \begin{bmatrix} \frac{1}{6} & \frac{1}{12} & \frac{1}{30} & \frac{1}{30} & \frac{1}{60} & 0 & 0 & 0 & \frac{1}{9} \\ \frac{1}{12} & \frac{6}{12} & \frac{1}{30} & \frac{60}{60} & \frac{30}{30} & 0 & 0 & 0 & \frac{1}{9} \\ \frac{1}{30} & \frac{1}{30} & \frac{90}{90} & \frac{180}{180} & \frac{180}{180} & 0 & 0 & 0 & \frac{1}{36} \\ \frac{1}{60} & \frac{60}{60} & \frac{180}{180} & \frac{90}{90} & \frac{180}{180} & 0 & 0 & 0 & \frac{1}{36} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{18} & \frac{1}{36} & \frac{1}{36} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{36} & \frac{1}{18} & \frac{1}{36} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{36}{36} & \frac{36}{36} & \frac{18}{18} & 0 \\ \frac{1}{9} & \frac{1}{9} & \frac{1}{36} & \frac{1}{36} & \frac{1}{36} & 0 & 0 & 0 & \frac{1}{5} \end{bmatrix}.$$

As with the D-optimality criterion, we define the Bayesian I-optimality criterion as the I-optimality criterion averaged over the prior distribution $\pi(\boldsymbol{\theta})$ of the parameter vector $\boldsymbol{\theta}$:

$$\mathcal{I}_B = \int_{\mathbb{R}^m} \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta}) \mathbf{W}] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (3.17)$$

3.3.4 Numerical approximation to optimality criteria

The Bayesian optimality criteria must be approximated numerically because there is no closed-form solution to the integrals in Equations (3.13) and (3.17). This is usually done by using random or systematic draws from the prior distribution $\pi(\boldsymbol{\theta})$ (Kessels et al., 2009; Ruseckaite et al., 2017; Train, 2009; Yu et al., 2010). In our work, we utilize Halton draws from the prior distribution because they reduce the variance of the approximation to the integral and provide a good coverage of the entire domain of the prior distribution (Train, 2009; Yu et al., 2010). Moreover, Bhat (2001) verified that around 100 Halton draws provide about the same level of accuracy as 2000 pseudo-random draws in the context of a 5-dimensional approximation to the likelihood of a mixed multinomial model. Yu et al. (2010) showed that Halton draws also produce good approximations of integrals with higher dimensions in the context of optimal design for choice experiments. Denoting the number of Halton draws by R and each individual draw by $\boldsymbol{\theta}^{(i)}$, our approximations for Equations (3.13) and (3.17) are

$$\mathcal{D}_B \approx \frac{1}{R} \sum_{i=1}^R \left[\det \left(\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta}^{(i)}) \right) \right]^{\frac{1}{m}}, \quad (3.18)$$

and

$$\mathcal{I}_B \approx \frac{1}{R} \sum_{i=1}^R \text{tr} \left[\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\theta}^{(i)}) \mathbf{W} \right], \quad (3.19)$$

respectively.

Like Ruseckaite et al. (2017) and in Chapter 2, we used $R = 128$ Halton draws from a multivariate normal prior distribution in both of our examples in the next section. We verified numerically that this number of draws provided a sufficiently good approximation of the Bayesian optimality criteria for the numbers of parameters in the models used in the two examples.

3.3.5 Construction of D- and I-optimal designs

To compute our optimal designs, we used a coordinate-exchange algorithm (Meyer and Nachtsheim, 1995; Goos and Jones, 2011). A coordinate-exchange algorithm was also used by Kessels et al. (2009), Ruseckaite et al. (2017), and in Chapter 2 in the context of choice experimentation. Our algorithm was implemented using the R programming language (R Core Team, 2017) with the aid of several existing R packages (Hamilton and Ferry, 2018; Wickham, 2016; Wickham et al., 2020; Eddelbuettel and François, 2011; Eddelbuettel, 2013;

Eddelbuettel and Balamuta, 2018; Eddelbuettel and Sanderson, 2014; Henry and Wickham, 2020), and created a package called `opdesmixr`, available at <https://github.com/mariobecerra/opdesmixr>, which allows the computation of locally D-optimal, Bayesian D-optimal, locally I-optimal, and Bayesian I-optimal designs for first-order, second-order, and special-cubic Scheffé models. We extended the package and added the functionality to compute locally D-optimal, Bayesian D-optimal, locally I-optimal and Bayesian I-optimal designs for the model presented in Equation (3.11), involving mixture ingredient proportions as well as process variables.

The coordinate-exchange algorithm we implemented starts from a random initial design, and begins by optimizing the first ingredient proportion of the first alternative within the first choice set, followed by the second ingredient proportion of the first alternative within the first choice set, and so on, until all q ingredient proportions have been optimized. Then, it continues with each of the r process variables. The algorithm then repeats this process for each alternative in each choice set in the design. The whole process is repeated until the design can no longer be improved or until a maximum number of iterations has been reached. At each step of the coordinate-exchange algorithm, we seek the optimal value of every individual ingredient proportion x_{ijs} or process variable setting z_{ijs} . This is a univariate optimization problem which can be solved in a straightforward way using Brent's univariate optimization method (Brent, 1973). Every time Brent's univariate optimization method is invoked during the course of the coordinate-exchange algorithm, the Bayesian D- or I-optimality criterion has to be evaluated. Despite the efficient approximation of these criteria using Halton draws, this renders the coordinate-exchange algorithm for choice experiments computationally intensive.

As indicated in Piepel et al. (2005), Goos and Jones (2011), Ruseckaite et al. (2017), and in Chapter 2, the coordinate-exchange algorithm must be modified to deal with mixtures. Since the mixture proportions must sum up to one, they cannot be independently changed. As a matter of fact, a change in one proportion requires a change in at least one other proportion. This dependency is solved by using the so-called Cox effect direction (Cornell, 2002; Goos and Jones, 2011; Piepel et al., 2005). After a change of one of the ingredient proportions, x_{ijs} , to $x_{ijs} + \Delta$, we modify the other $q - 1$ proportions as follows:

$$x_{kjs}^{\text{new}} = \begin{cases} \left(1 - \frac{\Delta}{1-x_{ijs}}\right) x_{kjs} & \text{if } x_{ijs} \neq 1, \\ \frac{1-(x_{ijs}+\Delta)}{q-1} & \text{if } x_{ijs} = 1. \end{cases}$$

3.4 Results

In this section, as proofs of concept, we present D- and I-optimal designs for two example choice experiments involving a mixture and one or more process variables. In both examples, we use a normal prior distribution, which is the most commonly used prior in the literature on the optimal design of choice experiments. The first example involves a cocktail tasting experiment and was inspired by Courcoux

and Séménou (1997b), while the second example involves a fish patty experiment. The inspiration for this example came from Cornell (1988, 2002) and Cornell and Gorman (1984) and Goos (2022).

3.4.1 Cocktail example

Courcoux and Séménou (1997b) discussed an experiment in which cocktails involving mango juice, blackcurrant syrup, and lemon juice were tasted. The experiment was conducted by asking respondents to taste different pairs of cocktails and indicating their preferred one in each pair. Ruseckaite et al. (2017) and us in Chapter 2 revisited this experiment, computed prior distributions for the parameter vector θ of a special-cubic Scheffé model, and created optimal experimental designs using this prior.

In the experiment, Courcoux and Séménou (1997b) imposed lower bounds of 0.3, 0.15 and 0.1 on the three ingredient proportions. To deal with this issue and to be able to use our implementation of the coordinate-exchange algorithm, like Ruseckaite et al. (2017) and like in Chapter 2, we expressed the mixtures defining the cocktails in terms of so-called pseudocomponents x_1 , x_2 , and x_3 . These pseudocomponents are defined such that they take a minimum value of 0 and a maximum value of 1, and sum up to one. The conversion of the true ingredient proportions into pseudocomponent proportions is done via the formula $x_i = (a_i - L_i)/(1 - L)$, where L_i denotes the lower bound of ingredient i , a_i denotes the true ingredient proportion, and L is the sum of the lower bounds for all q ingredient proportions.

In September 2019, students at KU Leuven replicated the experiment by asking 35 respondents to taste cocktails made with mango juice, blackcurrant syrup, and lemon juice and say which one they preferred. Each respondent tasted four choice sets of two cocktails. This experiment, as well as the original by Courcoux and Séménou (1997b), did not have any process variables. Due to the lack of a practical choice experiment that included both mixture and process variables, we decided to use the dataset from the KU Leuven students as an example of how to create a new optimal experiment. This situation reflects real-life scenarios in which experimenters have to plan and design an experiment using prior information, but they lack detailed information concerning the impact of at least certain variables. To use the KU Leuven data, we created additional simulated responses with a synthetic process variable which could be the cocktail's temperature to obtain a prior normal distribution for parameter vector θ using the model in Equation (3.11). We created these simulated responses with the assumption that responders would prefer a colder cocktail over a warmer one. We then fitted a multinomial logit model to these data, which gave us an estimated mean and variance-covariance-matrix, which in turn we used to construct the prior distribution in our cocktail example. Our prior mean vector is $\theta = (7.562, 0.907, 5.109, 14.573, 17.1806, 19.2705, 19.2705, 0)^T$. This means

that the utility of alternative j in choice set s was modeled as

$$\begin{aligned} U_{js} = & 7.562x_{1js} + 0.907x_{2js} \\ & + 5.109x_{1js}x_{2js} + 14.573x_{1js}x_{3js} + 17.1806x_{2js}x_{3js} \\ & + 19.2705x_{1js}z_{1js} + 19.2705x_{2js}z_{1js} + 19.2705x_{3js}z_{1js} \\ & + 0z_{1js}^2 + \varepsilon_{js}. \end{aligned}$$

The prior variance-covariance matrix we used is

$$\Sigma_0 = \text{diag}(4, 9, 49, 36, 49, 900, 900, 900, 900).$$

It must be noted that the variances in this matrix were rounded to the nearest integer from the estimated multinomial logit model. With this prior distribution, we computed Bayesian D- and I-optimal designs using the coordinate-exchange algorithm discussed in Section 3.3.5.

Our Bayesian D- and I-optimal designs are shown graphically in Figure 3.1. In the figure, the mixtures in each of the $35 \times 4 = 140$ choice sets are presented in terms of the pseudocomponent proportions. The shade of blue of each dot denotes the level of process variable temperature for the corresponding mixture. Figure 3.2 shows the distribution of the temperatures selected for the alternatives in the 140 choice sets in each of the designs.

In Figure 3.1, it can be seen that the points in the I-optimal design are spread more evenly over the entire simplex compared to those of the D-optimal counterpart. This is consistent with the results of Chapter 2 for choice experiments with mixtures in the absence of process variables. It is also worth pointing out that both designs use levels other than -1 and $+1$ for the process variable temperature, even though the mean prior value for the quadratic effect of the process variable temperature is zero.

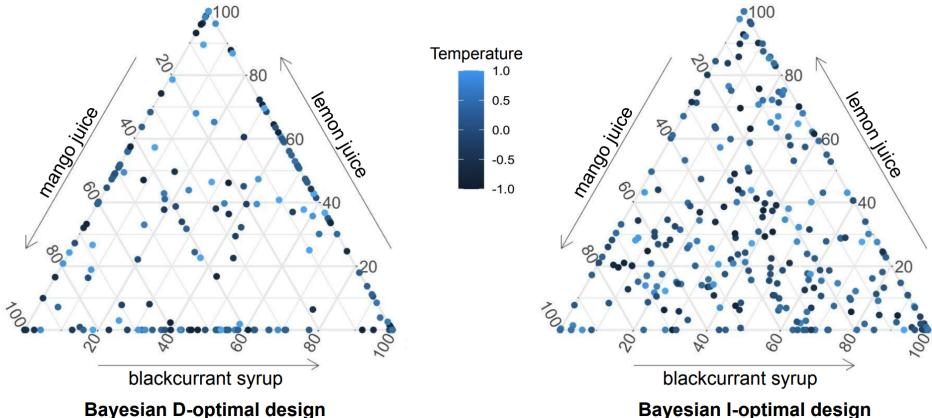


Figure 3.1: Bayesian D- and I-optimal designs produced by our coordinate-exchange algorithm for the cocktail experiment.

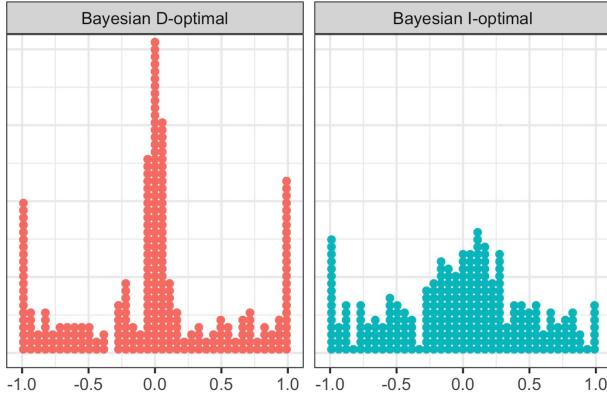


Figure 3.2: Distribution of the value of the process variable temperature in the Bayesian D- and I-optimal designs for the cocktail experiment.

Figure 3.3 shows the fraction of design space plots of the two Bayesian optimal designs. These plots display the performance of the designs in terms of the prediction variance for each point in the experimental region or design space (Zahran et al., 2003). The horizontal axis corresponds to a fraction of the experimental region, while the vertical axis ranges from the minimum to the maximum prediction variance over the entire experimental region (Goos & Jones, 2011). A curve in a fraction of design space plot shows the prediction variances $\mathbf{f}^T(\mathbf{x})\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta})\mathbf{f}(\mathbf{x})$ for a large number of random points selected from the experimental region, ordered from small to large. Ideally, all prediction variances are small throughout the entire experimental region, in which case the curve in the fraction of design space plot is virtually flat. Another way of explaining the fraction of design space plot is to say that it is the cumulative distribution function of the prediction variances across the experimental region, but with the positions of the two axes swapped.

The typical method to construct a fraction of design space plot for a given design is to randomly sample a large number of points M (e.g., 10,000 points) inside the experimental region. Then, the prediction variance $\mathbf{f}^T(\mathbf{x})\mathbf{I}^{-1}(\mathbf{X}, \boldsymbol{\beta})\mathbf{f}(\mathbf{x})$ is calculated for each of these points, and all M prediction variances are sorted from smallest to largest to obtain the empirical cumulative distribution function of the prediction variances (Ozol-Godfrey et al., 2005; Goldfarb et al., 2004; Goos and Jones, 2011). If we denote the prediction variance of the i -th sampled point by v_i , then the non-decreasing curve joining the M pairs $(i/M, v_i)$ forms the fraction of design space plot. A point i/M on the horizontal axis of the fraction of design space plot gives the proportion of the design space that has a prediction variance less than or equal to the corresponding value v_i on the vertical axis (Smucker et al., 2018). In order to deal with the issue of the prediction variance depending on the unknown parameter vector, we computed prediction variances for 128 Halton draws from the prior distribution of the parameter vector $\boldsymbol{\theta}$ and averaged the results.

The main takeaway from Figure 3.3 is that the prediction variance is much higher

for the Bayesian D-optimal design than for its I-optimal counterpart. The median prediction variance for the Bayesian D-optimal design is about 21.6, while it is about 10.9 for the Bayesian I-optimal one.

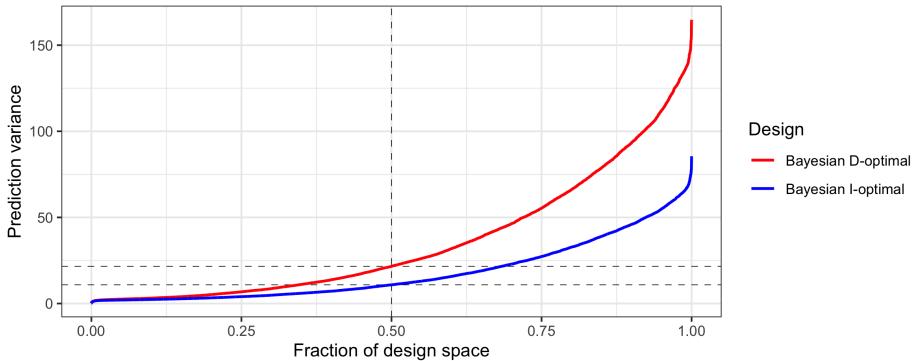


Figure 3.3: Fraction of design space plot of our Bayesian D- and I-optimal designs for the cocktail experiment.

3.4.2 Fish patty example

The second example we discuss involves a fish patty and was inspired by the work of Cornell (1988, 2002) and Cornell and Gorman (1984), Goos (2022). In the original experiment, the interest was in the firmness of patties made with a mixture of three fish species: mullet, sheepshead, and croaker. These patties were subjected to different processing conditions: oven cooking temperature (375 or 425 degrees Fahrenheit), oven cooking time (25 or 40 minutes), and deep fat frying time (25 or 40 seconds). The first three variables are mixture variables and the last three are process variables.

Since the original interest was in the firmness of the patty, no preference data is available to construct a normal prior distribution for our example. However, assuming firmness is proportional to utility, we used the original data and the model

$$\begin{aligned}
 Y = & \gamma_1^0 x_1 + \gamma_2^0 x_2 + \gamma_3^0 x_3 \\
 & + \gamma_{12}^0 x_1 x_2 + \gamma_{13}^0 x_1 x_3 + \gamma_{23}^0 x_2 x_3 \\
 & + \gamma_1^1 x_1 z_1 + \gamma_2^1 x_2 z_1 + \gamma_3^1 x_3 z_1 \\
 & + \gamma_1^2 x_1 z_2 + \gamma_2^2 x_2 z_2 + \gamma_3^2 x_3 z_2 \\
 & + \gamma_1^3 x_1 z_3 + \gamma_2^3 x_2 z_3 + \gamma_3^3 x_3 z_3 \\
 & + \alpha_{12} z_1 z_2 + \alpha_{13} z_1 z_3 + \alpha_{23} z_2 z_3 + \varepsilon
 \end{aligned}$$

to obtain a prior point estimate for the parameter vector $\boldsymbol{\theta}$. This model is the same as the one in Equation (3.8), but without the quadratic terms for the three

process variables. The reason we did not include these quadratic effects is that, in the original experiment, the process variables were studied at two levels only. As a consequence, the quadratic effects were inestimable. We obtained the following estimate for the parameter vector

$$\begin{aligned}\boldsymbol{\theta}^T &= (\gamma_1^0, \gamma_2^0, \gamma_3^0, \gamma_{12}^0, \gamma_{13}^0, \gamma_{23}^0, \gamma_1^1, \gamma_2^1, \gamma_3^1, \\ &\quad \gamma_1^2, \gamma_2^2, \gamma_3^2, \gamma_1^3, \gamma_2^3, \gamma_3^3, \alpha_{12}, \alpha_{13}, \alpha_{23}) \\ &= (2.864, 1.074, 2.003, -0.974, -0.834, 0.356, 0.376, 0.106, 0.206, \\ &\quad 0.642, 0.2, 0.403, -0.078, -0.087, -0.01, 0.027, 0.001, -0.008).\end{aligned}$$

Next, we transformed the parameter vector to the identified parameter space, as explained in Section 3.2.3. To this end, we computed $\gamma_1^{0*} = \gamma_1^0 - \gamma_3^0 = 2.864 - 2.003 = 0.861$ and $\gamma_2^{0*} = \gamma_2^0 - \gamma_3^0 = 1.074 - 2.003 = -0.929$. As a result, our prior model for the utility of alternative j in choice set s in the fish patty example is

$$\begin{aligned}U_{js} &= 0.861x_{1js} - 0.929x_{2js} \\ &\quad - 0.974x_{1js}x_{2js} - 0.834x_{1js}x_{3js} + 0.356x_{2js}x_{3js} \\ &\quad + 0.376x_{1js}z_{1js} + 0.106x_{2js}z_{1js} + 0.206x_{3js}z_{1js} \\ &\quad + 0.642x_{1js}z_{2js} + 0.2x_{2js}z_{2js} + 0.403x_{3js}z_{2js} \\ &\quad - 0.078x_{1js}z_{3js} - 0.087x_{2js}z_{3js} - 0.01x_{3js}z_{3js} \\ &\quad + 0.027z_{1js}z_{2js} + 0.001z_{1js}z_{3js} - 0.008z_{2js}z_{3js} \\ &\quad + 0z_{1js}^2 + 0z_{2js}^2 + 0z_{3js}^2 + \varepsilon_{js}.\end{aligned}$$

The estimates of the parameters in the initial model were used as the means of a set of normal prior distributions with variance-covariance matrices of the form $\boldsymbol{\Sigma}_0 = \kappa \mathbf{I}_{21}$, where κ is a positive scalar that controls the level of uncertainty and \mathbf{I}_{21} is the identity matrix of size 21. A higher value of κ indicates a higher level of uncertainty concerning the parameter values. This structure of variance-covariance gives us a simple way to study the impact of different levels of uncertainty expressed by the prior distribution on the final design.

The variance-covariance matrix $\boldsymbol{\Sigma}_0$ corresponding to the initial 21-parameter model must then also be transformed to the identified 20-dimensional parameter space. This results in a new 20×20 prior variance-covariance matrix

$$\boldsymbol{\Sigma}'_0 = \begin{pmatrix} 2\kappa & \kappa & 0 & \dots & 0 & 0 \\ \kappa & 2\kappa & 0 & \dots & 0 & 0 \\ 0 & 0 & \kappa & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \kappa & 0 \\ 0 & 0 & 0 & \dots & 0 & \kappa \end{pmatrix}.$$

We computed Bayesian D- and I-optimal designs for the same κ values as Ruseckaite et al. (2017) and Chapter 2, that is 0.5, 5, 10 and 30. All of our Bayesian D- and I-optimal designs are shown graphically in Figures 3.4 and 3.5. It can be seen that the spread in the points in the optimal designs increases with κ , and the spread is more pronounced for the Bayesian I-optimal designs than for the Bayesian D-optimal designs.

Figure 3.6 shows the fraction of design space plots for the Bayesian D- and I-optimal designs. For each value of κ , the D-optimal design has a much higher prediction variance than its I-optimal counterpart. Hence, the Bayesian I-optimal designs add substantial value in terms of precision of prediction when compared to Bayesian D-optimal designs.

3.5 Discussion

We introduced the theory for choice experiments involving mixtures and process variables, and embedded the Bayesian D- and I-optimality criteria in a coordinate-exchange algorithm for constructing optimal designs for this type of choice experiments. We also showed two examples in which the I-optimal designs perform substantially better than their D-optimal counterparts in terms of the variance of the predicted utility, which is something desirable because it is crucial to have precise predictions for any combination of ingredient proportions and process variables when optimizing the formulation of a mixture and the settings of the related process variables.

We would like to point out that the work we presented here has applications in other fields of research than food. This is because choice experiments involving mixtures are relevant in, for example, transportation and economics too. As a matter of fact, Zijlstra et al. (2019) conducted a choice experiment in which the mixtures between which the respondents had to choose were different ways in which a given mobility budget could be spent. Khademi et al. (2013) discussed a choice experiment involving a mixture of road toll, congestion pricing and parking price. Boonaert et al. (2023) used a choice experiment concerning the desired composition of a family, where the family composition is considered a mixture of boys and girls with different education levels. Finally, Yang et al. (2016) used a mixture choice experiment to measure context-dependent responses to accumulative energy charges under budget constraints. In all of these non-food-related choice experiments, an ad-hoc experimental design was used and there was a variable related to the total amount of the mixture. This total amount can be viewed as a process variable, and, therefore, the models and the optimal design approach we present here would be applicable to these choice experiments too.

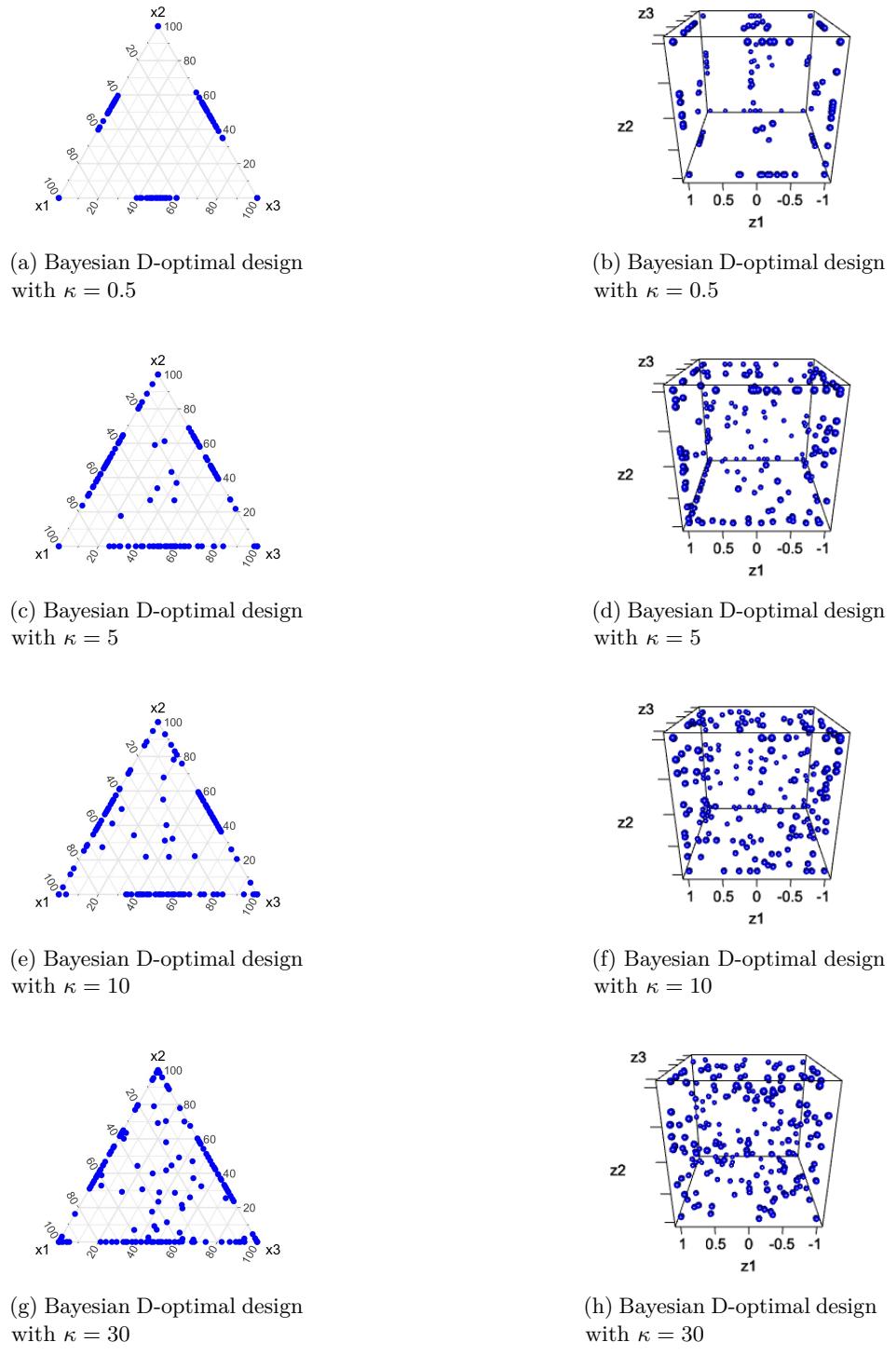


Figure 3.4: Bayesian D-optimal designs for the fish patty experiment. The four figures on the left show the mixture ingredient proportions, while the four figures on the right show the settings of the process variables.

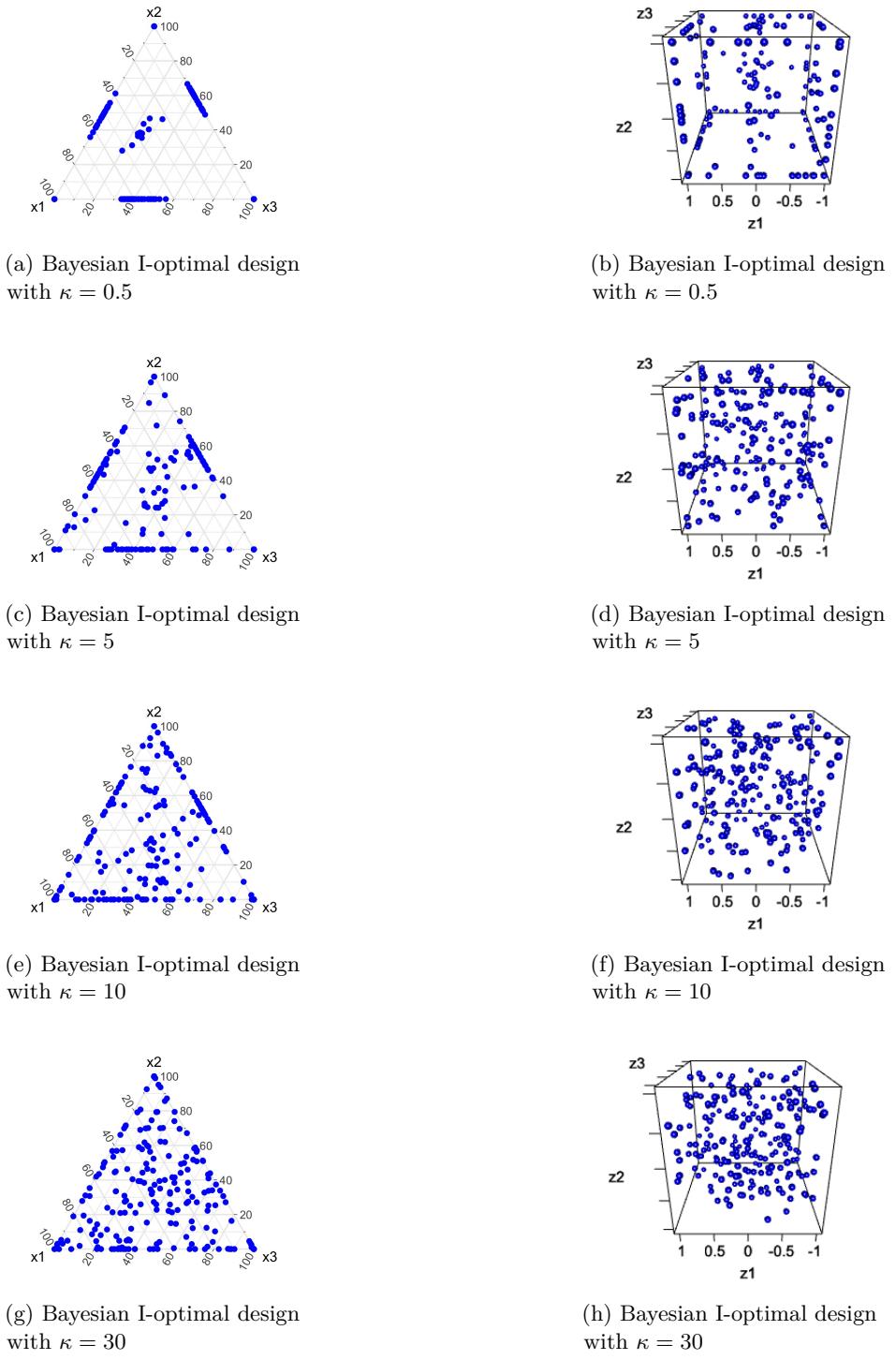


Figure 3.5: Bayesian I-optimal designs for the fish patty experiment. The four figures on the left show the mixture ingredient proportions, while the four figures on the right show the settings of the process variables.

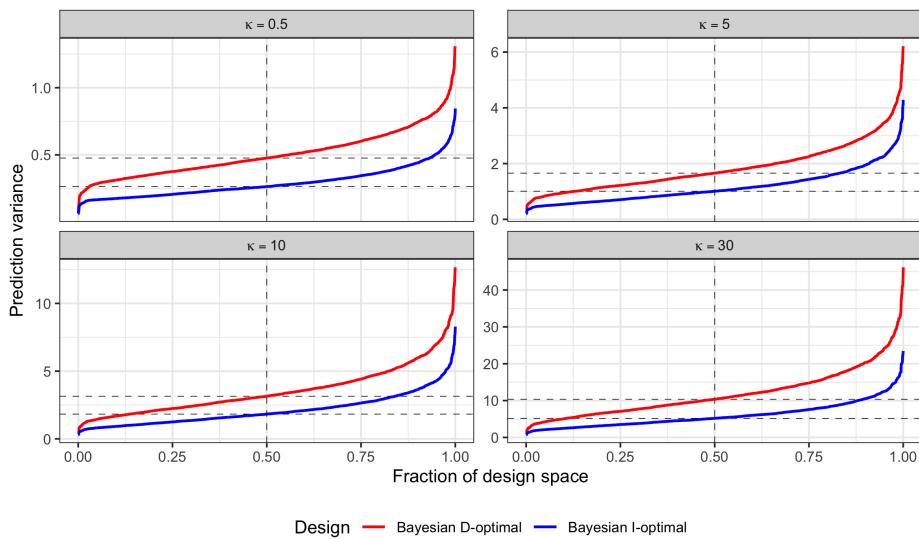


Figure 3.6: Fraction of design space plots of our Bayesian D- and I-optimal designs for the fish patty experiment for four values of κ , which represents the level of uncertainty concerning the prior parameter vector θ .

Chapter 4

Finding fruit flies' favorite color: a discrete choice experiment

Abstract

In recent years, a particular species of fruit fly originally from Southeast Asia, called *Drosophila suzukii*, has invaded Europe and has been causing damage to local agriculture, making it an economic threat. For this reason, methods that can trap and monitor these fruit flies are sought. Since the flies use visual and olfactory cues to find food, most traps are based on attraction using color or smell. Hence, studying the color preference of *Drosophila suzukii* is useful for the creation of traps. We therefore designed and carried out several choice experiments with said fruit flies in order to model their color preferences. We built a mixture-amount model of individual LEDs and their intensities to model the attractiveness of colors, and embedded that model in a mixed logit choice model. Using the model, we identified what the most attractive combinations of colors are, and we ran a confirmatory experiment. We found that *Drosophila suzukii* are attracted to UV light and that a combination of UV light and other colors is even more attractive. We conclude that our methodology was appropriate to identify the most attractive colors for the *Drosophila suzukii* and can be used to create traps.

4.1 Introduction

The common fruit fly (*Drosophila melanogaster*) is ubiquitous in orchards and homes all around the world. It is one of the most widely documented pest insects and it has been extensively studied. In the last decade, another species of fruit fly originally from Southeast Asia has invaded Europe, North America, South America, and northern Africa (Cruz-Esteban, 2021; Little et al., 2019; Mortelmanns et al., 2012; Tait et al., 2021). This species is the *Drosophila suzukii*, also known as the Asian vinegar fly or spotted wing Drosophila (SWD), and it has been causing damage to local agriculture, making it an economic threat (Tait et al., 2021; Walsh et al., 2011). This threat is due to the fact that, unlike the common fruit fly that likes overripe and rotting fruit, *Drosophila suzukii* prefers fruit that is still in the ripening process (Asplen et al., 2015; Hamby et al., 2013; Mitsui et al., 2006). For this reason, monitoring the presence of the species is of importance. Fruit flies use visual and olfactory cues to find food. Therefore, most traps are based on attraction using color or smell. As a result, it is useful to study the color preference of *Drosophila suzukii* in order to be able to create traps that will attract the flies.

Some authors have suggested that *Drosophila suzukii* have a preference for yellow colored traps (Iglesias et al., 2014; Lee et al., 2013). Others have suggested that they like red or black colors (Basoalto et al., 2013; Fountain et al., 2020; Lasa et al., 2017; Renkema et al., 2014; Rice et al., 2016). Additionally, Little et al. (2019) found that *Drosophila suzukii* seems to be attracted to ultraviolet (UV) light, and also suggested that contrast against the background might have a bigger effect than individual color appearance, and that the aforementioned attraction to red might be because of the contrast between darkness and brightness. Cruz-Esteban (2021) found that a yellow card inside a transparent trap with olfactory bait attracted a higher number of flies than other contrasting pairs of colors like red and black, green and purple, or yellow and green. This author argued that perhaps the flies were most attracted to the high reflectance offered by the color, as well as to the high luminosity of yellow compared to other colors. He concluded that reflectance of colors in contrast with the leaves' green color affects the capturing of *Drosophila suzukii*. In any case, there is no consensus in the literature about the color preference of the flies.

With this in mind, we study the color preference of *Drosophila suzukii* in a novel way. Most of the existing literature studies the color attractiveness by means of a categorical variable with predefined levels, e.g., red, yellow, black. Additionally, the response is usually modeled using linear regression, by either taking the raw number of flies trapped, or some sort of average as the response. The first novelty in our work is that we model the colors in a continuous way so that we can predict what color would be the most attractive to the flies. We did this by using a set of LEDs with 7 simple colors: red, orange, yellow, green, blue, purple, and ultraviolet, and defining the more complex colors as a mixture of these seven simple colors. In our model we also included an intensity variable related to how bright the overall color is. The second novelty is that we study the flies' preferences through a series of discrete choice experiments in which we show two colors to the flies, measure how many are attracted to each alternative, and use this measurement to fit a choice

model that can be used to predict the attractiveness of any color combination. This kind of modeling of preferences is standard in the choice modeling literature, but to our knowledge, it has not been used to model the preferences of fruit flies or other insects.

Choice modeling is frequently used to quantify preferences. This is often done through discrete choice experiments. Choice modeling and discrete choice experiments have been successfully applied in marketing (Rossi et al., 2012; Train, 2009), transportation (Khademi & Timmermans, 2012; Khademi et al., 2013; Zijlstra et al., 2019), ecology (Melero et al., 2018; Vardakakis et al., 2015), environmental economics (Bennett & Blamey, 2001; Torres et al., 2013; Vojáček, Pecáková, et al., 2010), sports (Balliauw et al., 2020; Becerra et al., 2023), and food research (Apellado-Buenaventura & Valmorida, 2021; Lizin et al., 2022; Skaltsi et al., 2022). Discrete choice experiments are carried out by presenting respondents sets of alternatives, called choice sets, and letting them choose between the alternatives. The respondents repeat this task several times with different choice sets. In the case of human respondents, this might be done in the form of surveys. With animals, it could be by monitoring their behavior in their natural habitat (Vardakakis et al., 2015) or by using traps (Melero et al., 2018). We were able to work in a controlled environment that allowed us to present different alternatives to fruit flies and count how many were attracted to each alternative.

On another note, modeling products, services and other items as mixtures is not uncommon in the statistical literature. Examples of mixture ingredients include the chemicals that are used to create a pesticide (Cornell, 2011); media used in advertising campaigns (Goos et al., 2019); components of a mobility budget (Zijlstra et al., 2019); cement, water, and sand to make concrete (Cornell, 2002); the wheat varieties used to bake bread (Rehman et al., 2007); and the ingredients used to make a drink (Courcoux & Séménou, 1997b; Goos & Hamidouche, 2019). In mixture experiments, the products, services and items under investigation are expressed as combinations of proportions of ingredients and the researchers' interest is generally in one or more characteristics of the mixture. As mentioned before, the characteristic of interest in our work is the color preference of *Drosophila suzukii*. Therefore, using mixtures and choice experiments is a reasonable approach to solve our problem at hand.

Despite the fact that numerous products, services and other items involve mixtures of ingredients, literature concerning choice experiments with mixtures is scarce. Some of the few works on this topic include the work by Courcoux and Séménou (1997b), who performed a taste experiment to model the preferences for cocktails involving different proportions of mango juice, lime juice, and blackcurrant syrup. Goos and Hamidouche (2019) used the data from Courcoux and Séménou (1997b) and explained how to embed Scheffé models for data from mixture experiments in different logit type models for choice experiments. Khademi and Timmermans (2012) analyzed the response of travelers to mixtures of transport charges subject to budget constraints. Khademi et al. (2013) discussed a choice experiment involving a mixture of road toll, congestion pricing, and parking price. Yang et al. (2016) used a mixture choice experiment to measure context-dependent responses to accumulative energy charges under budget constraints. Zijlstra

et al. (2019) performed a choice experiment to investigate the preferred mobility budget mixture for employees of a company. Boonaert et al. (2023) used a choice experiment concerning the desired composition of a family in sub-Saharan Africa, where the family composition was considered a mixture of boys and girls with different education levels. In the context of experimental design, Ruseckaite et al. (2017) compared two algorithms to find D-optimal designs for choice experiments with mixtures assuming a multinomial logit model. Building on this work, in Chapter 2 we constructed I-optimal designs for choice experiments with mixtures. In Chapter 3 we extended this work and included additional variables unrelated to the composition of the mixture. These additional variables are important when the preference for a mixture may depend on characteristics other than its composition, such as the temperature at which a cocktail is served, or the baking time and the baking temperature used for a bread. Variables describing such characteristics are typically called *process variables* (Goos & Jones, 2011).

One particular case of a process variable is the total amount of the mixture ingredients. An example of such scenario can be found in Zijlstra et al. (2019), who observed that the preferred mobility budget mixture depends on the budget amount. Another example is Boonaert et al. (2023), who confirmed the existence of a quantity-quality trade-off in Ethiopia between the family size and level of education of their children. In the general literature on mixture experiments, models that involve mixture ingredients and the total amount are typically called *mixture-amount* models (Piepel & Cornell, 1987). In this chapter, the amount variable is the total intensity of the light shown to the fruit flies.

The rest of the chapter is organized as follows. In Section 4.2, we introduce the most often used models for mixture-amount experiments and the multinomial logit model for choice data, and we explain how to embed the mixture-amount models in the multinomial logit model. We also introduce the mixture-amount mixed logit model which we use for the fruit fly data. In Section 4.3, we discuss I-optimality, one of the most commonly used metrics to measure the quality of experimental designs, and which we used to design our experiments. In Section 4.4, we show the experimental setup of our experiments. In Section 4.5, we present our results and provide insights about *Drosophila suzukii*'s color preference. Finally, in Section 4.6, we summarize our work.

4.2 Models

In this section, we introduce the most commonly used models for data from mixture-amount experiments and the multinomial logit model for choice data, and explain how to embed the mixture-amount models in the multinomial logit model. Finally, we give details about the mixture-amount mixed logit model we used to fit the data and make predictions about the preferences of the fruit flies.

4.2.1 Models for data from mixture-amount experiments

Ordinary mixture experiments involve two or more ingredients and a response variable that depends only on the relative proportions of the ingredients in the mixture. We denote the number of ingredients by q . Each mixture in the experiment is described as a combination of q ingredient proportions, with the constraint that these proportions sum up to one. Due to this constraint, a classical regression model involving an intercept and linear terms in the ingredient proportions exhibits perfect collinearity. Therefore, researchers must use dedicated regression models when analyzing data from mixture experiments. The most commonly used family of models for data from mixture experiments is the Scheffé family (Scheffé, 1958, 1963). The most popular Scheffé models are the first-order, second-order, and special-cubic models.

Denoting the response in a traditional mixture experiment with a continuous outcome by Y and the q ingredient proportions by x_1, x_2, \dots, x_q , with $x_i \geq 0$ and $\sum_{i=1}^q x_i = 1$, the first-order Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \varepsilon. \quad (4.1)$$

The second-order Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \varepsilon, \quad (4.2)$$

and, finally, the special-cubic Scheffé model is

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \sum_{i=1}^{q-2} \sum_{j=i+1}^{q-1} \sum_{k=j+1}^q \beta_{ijk} x_i x_j x_k + \varepsilon. \quad (4.3)$$

In all three models, ε denotes the error term, which, for continuous outcomes, is typically assumed to be normally distributed.

In addition to the q ingredient proportions, researchers might also be interested in the effect that the total amount of ingredients might have on the response variable. Models that involve q mixture ingredients and the total amount of such ingredients are usually called *mixture-amount* models (Piepel & Cornell, 1987). They can be obtained by combining Scheffé models for the ingredient proportions with a polynomial model for the amount variable. For example, consider the model by Kowalski et al. (2000), which was originally created for mixture ingredients and process variables. Since the total amount can be considered a special case of a process variable, we can adapt the model to be used with q mixture variables x_1, x_2, \dots, x_q and an amount variable z . The resulting model is as follows

$$Y = \sum_{i=1}^q \gamma_i^0 x_i + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_i x_k + \sum_{i=1}^q \gamma_i x_i z + \alpha z^2 + \varepsilon. \quad (4.4)$$

We use this model as our starting point for computing optimal designs for mixture-amount choice experiments, as well as for the modeling of the data obtained from these experiments.

4.2.2 Multinomial logit model for choice data

The multinomial logit model builds on random-utility theory and assumes that a respondent in a choice experiment evaluates S choice sets, each involving J alternatives. The model assumes that, within each choice set $s \in \{1, \dots, S\}$, each respondent chooses the alternative that has the highest perceived utility. Therefore, the probability that a respondent chooses alternative $j \in \{1, \dots, J\}$ in choice set s , denoted by p_{js} , is the probability that the perceived utility of alternative j in choice set s , denoted by U_{js} , is larger than that of the other alternatives in the choice set:

$$p_{js} = \mathbb{P}[U_{js} > \max(U_{1s}, \dots, U_{j-1,s}, U_{j+1,s}, \dots, U_{Js})].$$

Since, generally, each alternative in a choice set has a set of observable attributes that characterize it, the perceived utility U_{js} can be expressed as

$$U_{js} = \mathbf{f}^T(\mathbf{a}_{js})\boldsymbol{\beta} + \varepsilon_{js}, \quad (4.5)$$

where \mathbf{a}_{js} is the vector that contains the levels of the attributes corresponding to alternative j in choice set s , $\mathbf{f}(\mathbf{a}_{js})$ represents the model expansion of this attribute level vector, and $\boldsymbol{\beta}$ is the vector containing the model parameters. The model parameters contained within $\boldsymbol{\beta}$ express the preferences of the respondents for the alternatives' attribute levels. In the multinomial logit model, the error terms ε_{js} are assumed to be independent and identically Gumbel distributed. The Gumbel distribution is also known as the generalized extreme value distribution of type I and as the log-Weibull distribution. As a result of the distributional assumption, it can be shown that

$$p_{js} = \frac{\exp[\mathbf{f}^T(\mathbf{a}_{js})\boldsymbol{\beta}]}{\sum_{t=1}^J \exp[\mathbf{f}^T(\mathbf{a}_{ts})\boldsymbol{\beta}]} \quad (4.6)$$

4.2.3 Model for choice data concerning mixture-amount experiments

In this chapter, we focus on choice experiments involving mixtures with an amount variable. Therefore, we assume that the attributes of the alternatives in the choice sets are the proportions of the ingredients of a mixture and the total amount of the ingredients. Consequently, we assume that the attribute vector \mathbf{a}_{js} in Equations (4.5) and (4.6) contains the q ingredient proportions x_1, x_2, \dots, x_q and the amount variable z of the j -th alternative in choice set s , and that $\mathbf{f}(\mathbf{a}_{js})$ represents the model expansion of the proportions and the amount. We base the polynomial expansion $\mathbf{f}(\mathbf{a}_{js})$ on a model combining a second-order Scheffé model for the q ingredients in the mixture with a quadratic regression model for the amount variable, as in Equation (4.4).

When starting from the model in Equation (4.4), the most natural thing to do would be to write the perceived utility U_{js} of alternative j in choice set s as

$$U_{js} = \sum_{i=1}^q \gamma_i^0 x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} + \sum_{i=1}^q \gamma_i x_{ijs} z_{js} + \alpha z_{js}^2 + \varepsilon_{js},$$

where x_{ijs} denotes the proportion of the i -th mixture ingredient in alternative j from choice set s , z_{js} denotes the amount for alternative j in choice set s , and ε_{js} denotes the error term. However, as explained by Ruseckaite et al. (2017), Goos and Hamidouche (2019), and Becerra and Goos (2021, 2023), due to the constraint that the ingredient proportions sum up to one, this leads to an inestimable multinomial logit model. As a consequence of the constraint, we can rewrite x_{qjs} as $1 - x_{1js} - \dots - x_{q-1,js}$ and U_{js} as

$$\begin{aligned} U_{js} &= \sum_{i=1}^{q-1} \gamma_i^0 x_{ijs} + \gamma_q^0 (1 - x_{1js} - \dots - x_{q-1,js}) + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^q \gamma_i x_{ijs} z_{js} + \alpha z_{js}^2 + \varepsilon_{js} \\ &= \gamma_q^0 + \sum_{i=1}^{q-1} (\gamma_i^0 - \gamma_q^0) x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} \\ &\quad + \sum_{i=1}^q \gamma_i x_{ijs} z_{js} + \alpha z_{js}^2 + \varepsilon_{js}. \end{aligned}$$

The latter expression for the perceived utility U_{js} involves a constant, γ_q^0 . Since the multinomial logit model only takes into account differences in utility, that constant causes the model to be ill-defined and, hence, inestimable. This can be circumvented by dropping γ_q^0 . Defining the parameters $\gamma_i^{0*} = \gamma_i^0 - \gamma_q^0$ for $i \in \{1, \dots, q-1\}$, we obtain the following expression for the perceived utility:

$$U_{js} = \sum_{i=1}^{q-1} \gamma_i^{0*} x_{ijs} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ik}^0 x_{ijs} x_{kjs} + \sum_{i=1}^q \gamma_i x_{ijs} z_{js} + \alpha z_{js}^2 + \varepsilon_{js}. \quad (4.7)$$

The parameter vector β of the expression in Equation (4.7) then becomes

$$\beta = (\gamma_1^{0*}, \gamma_2^{0*}, \dots, \gamma_{q-1}^{0*}, \gamma_{1,2}^0, \dots, \gamma_{q-1,q}^0, \gamma_1, \gamma_2, \dots, \gamma_q, \alpha)^T,$$

which does lead to a well-defined, estimable model. This vector has $K = 2q + \frac{q(q-1)}{2}$ elements. Due to the omission of γ_q^0 , the interpretation of the main effect of the i -th proportion, γ_i^{0*} , must be done in relation to the main effect of the q -th proportion.

4.2.4 Mixed logit model

The multinomial logit model presented in Section 4.2.3 assumes that the preferences of all respondents do not vary across experiments. That is, there is only one

parameter vector for the entire population of respondents. As previously mentioned, we performed several experiments, and each experiment was performed with a different set of flies on different days and at different times of the day. Therefore, there is reason to think that there will be variability between the experiments. A mixed logit model is a good way to model this preference heterogeneity: instead of a single parameter vector for all the data, it has a set of parameters for each experiment, allowing to model the variability between the experiments (McFadden & Train, 2000; Train, 2009).

Our model assumes that, in each performed experiment m , each respondent (i.e. each fly) faces S choice sets involving J alternatives each, and that, within each choice set $s \in \{1, \dots, S\}$, each respondent chooses the alternative that has the highest perceived utility. We denote the probability that a respondent chooses alternative j in choice set s in experiment m by p_{jsm} and the perceived utility of alternative j in choice set s in experiment m by U_{jsm} , with $j \in \{1, \dots, J\}$, $s \in \{1, \dots, S\}$, and $m \in \{1, \dots, M\}$, where M is the total number of experiments. Our model assumes that p_{jsm} is the probability that the perceived utility U_{jsm} is larger than the perceived utilities of the other alternatives in the choice set. Since each alternative in a choice set is a combination of ingredient proportions and of the total amount, U_{jsm} can be expressed as

$$\begin{aligned} U_{jsm} &= \mathbf{f}^T(\mathbf{a}_{jsm})\boldsymbol{\beta}_m + \varepsilon_{jsm} \\ &= \sum_{i=1}^{q-1} \gamma_{im}^{0*} x_{ijsm} + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ikm}^0 x_{ijsm} x_{kjsm} + \sum_{k=1}^q \gamma_{km} x_{kjsm} z_{jsm} \\ &\quad + \alpha_m z_{jsm}^2 + \varepsilon_{jsm}, \end{aligned} \quad (4.8)$$

where \mathbf{a}_{jsm} is the attribute vector that contains the mixture proportions and total amount corresponding to alternative j in choice set s for experiment m , x_{kjsm} is the k -th ingredient proportion of alternative j in choice set s for experiment m , z_{jsm} is the value of the amount variable for alternative j in choice set s for experiment m , and $\boldsymbol{\beta}_m$ is the parameter vector of experiment m . Notice that Equation (4.8) is analogous to Equation (4.7) but with an extra index referring to the experiment for the parameters of the mixture variables x_{kjsm} , the amount variable z_{jsm} , and the error term.

The error terms ε_{jsm} are assumed to be independent and identically Gumbel distributed, which results in the following expression for the probability that a respondent chooses alternative j in choice set s in experiment m :

$$p_{jsm} = \frac{\exp(U_{jsm})}{\sum_{t=1}^J \exp(U_{tsm})}. \quad (4.9)$$

Because we expect preference heterogeneity across the experiments, in our mixed logit model, each experiment m has its own set of individual parameters $\boldsymbol{\beta}_m$. We assume that each $\boldsymbol{\beta}_m$ comes from a single population. More specifically, we assume that each $\boldsymbol{\beta}_m$ is drawn from a multivariate normal distribution with a fixed mean

and variance-covariance matrix:

$$\boldsymbol{\beta}_m \sim \text{MVN}(\boldsymbol{\beta}, \Sigma), \quad (4.10)$$

where MVN denotes a multivariate normal distribution, $\boldsymbol{\beta}$ is a vector of parameters of size K , and Σ is the $K \times K$ covariance matrix of the heterogeneity distribution.

One last detail about the model is the likelihood function. If we denote by y_{jsm} the number of flies choosing alternative j in choice set s in experiment m , then the log-likelihood is

$$LL(\boldsymbol{\theta}) = \sum_{m=1}^M \sum_{s=1}^S \sum_{j=1}^J y_{jsm} \log(p_{jsm}), \quad (4.11)$$

with p_{jsm} the choice probability defined in Equation (4.9) and $\boldsymbol{\theta}$ denoting the vector containing all the parameters in the model, namely $\boldsymbol{\beta}$, $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_M$, and Σ .

4.2.5 No-choice option

In our experiments, the flies have the option to not choose any of the colors presented, since they are in an enclosed box and can roam freely in it. To deal with this, our mixed-logit model had to be adapted: every choice set essentially involves an alternative $J + 1$, on top of the initial J alternatives. That extra alternative is called a *no-choice* or *opt-out* option. In the choice modeling literature, the inclusion of a no-choice or opt-out alternative is not uncommon (Goos et al., 2010; Haaijer et al., 2001; Johnson & Orme, 1996; Lanzsar & Louviere, 2008; Vermeulen et al., 2008).

We build on the so-called extended no-choice multinomial logit model in which an additional indicator variable is added to the utility equation (Vermeulen et al., 2008). This dummy variable takes a value of 1 if the no-choice option is selected and a value of 0 otherwise. The utility of the no-choice option is determined solely by the coefficient of the no-choice variable. This is the simplest way to add a no-choice alternative to our model as it only requires the addition of an extra variable in the utility equation. More complex models are possible and are left for future work.

Combining the extended no-choice multinomial logit model with our mixture-amount mixed logit model, the perceived utility of alternative j in choice set s in experiment m is modeled as

$$U_{jsm} = \begin{cases} \sum_{i=1}^{q-1} \gamma_{im}^{0*} x_{ijsm} \\ + \sum_{i=1}^{q-1} \sum_{k=i+1}^q \gamma_{ikm}^0 x_{ijsm} x_{kjsm} \\ + \sum_{i=1}^q \gamma_{im} x_{ijsm} z_{jsm} + \alpha_m z_{jsm}^2 + \varepsilon_{jsm}, & \text{for } j \in \{1, \dots, J\}, \\ \lambda_m + \varepsilon_{jsm} & \text{for the no-choice option,} \end{cases} \quad (4.12)$$

where λ_m is the parameter for the no-choice option. Hence, the parameter vector for each experiment m becomes

$$\boldsymbol{\beta}_m = (\gamma_{1,m}^{0*}, \gamma_{2,m}^{0*}, \dots, \gamma_{q-1,m}^{0*}, \gamma_{1,2,m}^0, \dots, \gamma_{q-1,q,m}^0, \gamma_{1,m}, \gamma_{2,m}, \dots, \gamma_{q,m}, \alpha_m, \lambda_m)^T,$$

the parameter vector $\boldsymbol{\beta}_m$ has an extra element corresponding to the no-choice option and Σ is now of dimension $(K + 1) \times (K + 1)$.

4.2.6 Estimation

We used a Bayesian hierarchical estimation approach to fit our mixed logit model (Allenby et al., 2005; Gelman et al., 2013; Train, 2001, 2009). Usually, the use of maximum-likelihood estimation for the mixed logit model involves the simulation of the likelihood function, which makes the maximization procedure computationally intensive in contrast to computing the likelihood analytically. In addition, the choice of initial parameter values is critical to avoid convergence failures (Train, 2009). The use of a Bayesian hierarchical approach helps avoid the problem of choosing initial parameter values and, even though the computation process is intensive, it allows for each of the experiment-level parameters $\boldsymbol{\beta}_m$ to borrow strength from each other, since the model does not estimate each experiment's parameters in a vacuum, but rather it uses information from all of the data (Gelman et al., 2013). Moreover, the use of Bayesian modeling has been standard in the choice modeling literature for many years now (Feit et al., 2017; Otter et al., 2022; Rossi, 2019; Rossi et al., 2012; Train, 2009).

For the Bayesian estimation, we must place prior distributions on the parameters $\boldsymbol{\beta}$ and Σ . In the case of $\boldsymbol{\beta}$, we assume a multivariate normal distribution

$$\boldsymbol{\beta} \sim \text{MVN}(\boldsymbol{\beta}^*, \Sigma^*), \quad (4.13)$$

where $\boldsymbol{\beta}^*$ and Σ^* are prior values for the mean and variance-covariance matrix respectively. As for Σ , it used to be standard to use an inverse-Wishart prior for this matrix because of desirable conjugacy properties. However, it has been shown that this prior in general might have undesirable properties (Akinc & Vandebroek, 2018; Álvarez et al., 2014; Liu et al., 2016; Tokuda et al., 2011). An alternative is to first decompose Σ into two elements: a scale vector and a correlation matrix. Particularly,

$$\Sigma = \text{diag}(\boldsymbol{\tau}) \times \Omega \times \text{diag}(\boldsymbol{\tau}), \quad (4.14)$$

where $\text{diag}(\boldsymbol{\tau})$ denotes a diagonal matrix of size $(K + 1) \times (K + 1)$ with entries equal to the elements of the scale vector $\boldsymbol{\tau}$, and Ω denotes a correlation matrix of size $(K + 1) \times (K + 1)$ (Barnard et al., 2000; Liu et al., 2016; McElreath, 2018; Stan Development Team, 2022). The parameters $\boldsymbol{\tau}$ and Ω have their own prior distributions. In particular, for each element of the $\boldsymbol{\tau}$ vector, our prior distribution is a univariate normal distribution

$$\tau_k \sim N(\mu^*, \sigma^{2*}), \quad (4.15)$$

where τ_k denotes the k -th element of $\boldsymbol{\tau}$, for $k \in \{1, \dots, K + 1\}$. As for the prior distribution of the Ω matrix, we use

$$\Omega \sim \text{LKJCorr}(n^*), \quad (4.16)$$

where $\text{LKJCorr}(\eta^*)$ denotes a LKJ correlation prior with shape parameter η^* (Lewandowski et al., 2009). A value of $\eta^* < 1$ favors more correlation between the elements of the β parameter vector, whereas $\eta^* > 1$ favors less correlation (Gelman et al., 2013; Stan Development Team, 2022).

In a fully Bayesian inference process, we obtain the posterior distribution of θ by using Bayes' theorem:

$$p(\theta|y, X) = \frac{p(y|\theta, X)p(\theta|X)}{p(y|X)}, \quad (4.17)$$

where y denotes the vector of responses (the number of flies for each alternative and the no-choice option in each choice set in each experiment), X denotes the design matrix, and $(y|\theta, X)$ denotes the multinomial likelihood, which was defined in its logarithmic form in Equation (4.11). Again, θ denotes the vector containing all the parameters in the model, namely β , β_1, \dots, β_M , τ , and Ω ; with these two latter being the basis for Σ . Equation (4.17) can be made more precise by using some definitions like conditional probability, the properties of independence between certain parameters, and the fact that y is affected by β , τ and Ω only through β_1, \dots, β_M . This results in the following:

$$\begin{aligned} p(\theta|y, X) &= p(\beta_1, \dots, \beta_M, \beta, \tau, \Omega|y, X) \\ &= \frac{p(y|\beta_1, \dots, \beta_M, \beta, \tau, \Omega, X)p(\beta_1, \dots, \beta_M, \beta, \tau, \Omega|X)}{p(y|X)} \\ &= \frac{p(y|\beta_1, \dots, \beta_M, X)p(\beta_1, \dots, \beta_M|\beta, \tau, \Omega, X)p(\beta)p(\tau)p(\Omega)}{p(y|X)}. \end{aligned} \quad (4.18)$$

4.3 I-optimal designs

In the literature on the optimal design of choice experiments, several criteria for selecting a highly informative experimental design have been studied (Kessels et al., 2006). However, in the literature on optimal design of choice experiments with mixtures, the only two optimality metrics that have been studied are D-optimality and I-optimality. In Chapter 3, we described in detail how to generate D- and I-optimal designs for choice experiments with mixtures involving process variables assuming a multinomial logit model. In this section, we briefly explain how to create I-optimal designs for our fruit fly experiments. Additionally, we show how to handle the no-choice option and how to include it in the definition of the I-optimality criterion.

The I-optimality criterion is a prediction-oriented criterion because it focuses on getting precise predictions with the estimated statistical model. It is generally defined as the average prediction variance over the experimental region. The I-optimality criterion is also sometimes called the V-optimality criterion (Goos & Syafitri, 2014; Kessels et al., 2006).

To create an I-optimal experimental design, we start from the information matrix corresponding to the model used. In the design of our experiments, we assumed a simple multinomial logit model, which facilitated the computations enormously. Under the multinomial logit model assumption, the information matrix depends on the unknown parameter vector β through the choice probabilities p_{js} defined in Equation (4.6). This dependence of the optimal design on the parameter vector is typical for models that are not linear in the parameters, such as discrete choice models, and it implies that prior information is needed to find optimal designs (Kessels et al., 2006). This information can be provided in the form of a point estimate, or in the form of a prior distribution (Atkinson & Haines, 1996; Becerra & Goos, 2021, 2023; Kessels et al., 2006; Ruseckaite et al., 2017). We used a prior distribution, which leads to so-called Bayesian optimal designs. Bayesian optimal designs take into account the uncertainty about the parameter vector β through the use of a prior distribution $\pi(\beta)$ that summarizes the prior knowledge concerning the parameter vector β .

The information matrix $\mathbf{I}(\mathbf{X}, \beta)$ for the multinomial logit model is the sum of the information matrices of each of the S choice sets (Kessels et al., 2006). In the absence of a no-choice option, the information matrix for the multinomial logit model is

$$\mathbf{I}(\mathbf{X}, \beta) = \sum_{s=1}^S \mathbf{X}_s^T (\mathbf{P}_s - \mathbf{p}_s \mathbf{p}_s^T) \mathbf{X}_s,$$

with $\mathbf{p}_s = (p_{1s}, \dots, p_{Js})^T$, $\mathbf{P}_s = \text{diag}(\mathbf{p}_s)$, $\mathbf{X}_s^T = [\mathbf{f}(\mathbf{a}_{1s}), \mathbf{f}(\mathbf{a}_{2s}), \dots, \mathbf{f}(\mathbf{a}_{Js})]$ the model matrix containing the model expansions of the attribute levels of all J alternatives in choice set s , and $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_S]$ the model matrix for all S choice sets. The inverse of the information matrix is the asymptotic variance-covariance matrix of the maximum likelihood estimates of the parameter vector β . The information matrix depends on β through the choice probabilities in \mathbf{p}_s and \mathbf{P}_s .

When using choice models, there are two ways in which we can define I-optimality. If the goal is to predict choice probabilities, the I-optimality criterion is the average variance of the predicted choice probabilities. If the goal is to predict utilities, as in our study with the fruit flies, the I-optimality criterion is the average variance of the predicted utilities. In Chapter 2, we introduced a computationally efficient definition for I-optimal designs for choice experiments focused on the predicted utilities, and in Chapter 3 we extended this definition to include process variables. This is the definition we will use here too. Under this definition, the Bayesian I-optimality criterion is

$$\mathcal{I}_B = \int_{\mathbb{R}^m} \text{tr} [\mathbf{I}^{-1}(\mathbf{X}, \beta) \mathbf{W}] \pi(\beta) d\beta. \quad (4.19)$$

where $\mathbf{I}^{-1}(\mathbf{X}, \beta)$ is the inverse of the information matrix for model matrix \mathbf{X} and prior parameter vector β , and $\pi(\beta)$ denotes the prior distribution of β . Assuming that the amount variable is rescaled to the $[-1, +1]$ interval, the matrix \mathbf{W} is the moments matrix, defined as

$$\mathbf{W} = \int_{\chi} \mathbf{f}(\mathbf{a}_{js}) \mathbf{f}^T(\mathbf{a}_{js}) d\mathbf{a}_{js}, \quad (4.20)$$

with χ the experimental region which, for a mixture-amount choice model, combines the $(q-1)$ -dimensional simplex S_{q-1} for the ingredient proportions and the $[-1, +1]$ interval for the amount variable.

To compute the moments matrix \mathbf{W} for the mixture-amount model described in Equation (4.7), we first need to compute the matrix $\mathbf{f}(\mathbf{a}_{js})\mathbf{f}^T(\mathbf{a}_{js})$. This matrix has elements of the form

$$z^m \prod_{k=1}^q x_k^{n_k},$$

for some $n_k, m \in \mathbb{N}$, $k \in \{1, \dots, q\}$. Hence, each element of the moments matrix \mathbf{W} is of the form

$$\int_{\chi} z^m \prod_{k=1}^q x_k^{n_k} dz dx_1 \dots dx_q = \left(\frac{1^{m+1} - (-1)^{m+1}}{m+1} \right) \left(\frac{\prod_{k=1}^q n_k!}{(q-1 + \sum_{k=1}^q n_k)!} \right).$$

If m is odd, then $1^{m+1} - (-1)^{m+1}$ is zero and hence the corresponding element of the moments matrix also becomes zero. If m is even, then $1^{m+1} - (-1)^{m+1}$ is equal to 2 and the corresponding element of the moments matrix is nonzero.

In this work, we have a no-choice option in every choice set. If we denote by \mathbf{W} the moments matrix for the model without the no-choice option, then the moments matrix for the model including the no-choice option is simply

$$\mathbf{W}_{nc} = \begin{bmatrix} \frac{J}{J+1}\mathbf{W} & \mathbf{0} \\ \mathbf{0}^T & \frac{2}{(J+1)(q-1)!} \end{bmatrix}, \quad (4.21)$$

where $\mathbf{0}$ denotes a K -dimensional column vector of zeros, with $K = 2q + \frac{q(q-1)}{2} - 1$; and J is the number of alternatives other than the no-choice option.

Because of the no-choice option, the design matrix also changes and, in turn, so does the information matrix. These changes are straightforward and come from the definition of the no-choice option. The information matrix for the multinomial logit model with the no-choice option is

$$\mathbf{I}_{nc}(\mathbf{X}_{nc}, \boldsymbol{\beta}) = \sum_{s=1}^S \mathbf{X}_{nc,s}^T (\mathbf{P}_{nc,s} - \mathbf{p}_{nc,s} \mathbf{p}_{nc,s}^T) \mathbf{X}_{nc,s},$$

with $\mathbf{X}_{nc,s}^T = [\mathbf{f}(\mathbf{a}_{1s}), \mathbf{f}(\mathbf{a}_{2s}), \dots, \mathbf{f}(\mathbf{a}_{(J+1)s})]$ the model matrix containing the model expansions of the attribute levels of all J alternatives and the $J+1$ no-choice option in choice set s , $\mathbf{p}_{nc,s} = (p_{1s}, \dots, p_{(J+1)s})^T$, $\mathbf{P}_{nc,s} = \text{diag}(\mathbf{p}_{nc,s})$, and $\mathbf{X}_{nc} = [\mathbf{X}_{nc,1}, \dots, \mathbf{X}_{nc,S}]$ the model matrix for all S choice sets.

The definition of I-optimality for the multinomial logit model with the no-choice option is therefore

$$\mathcal{I}_B = \int_{\mathbb{R}^m} \text{tr} [\mathbf{I}_{nc}^{-1}(\mathbf{X}_{nc}, \boldsymbol{\beta}) \mathbf{W}_{nc}] \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}. \quad (4.22)$$

The Bayesian I-optimality criterion must be approximated numerically because there is no closed-form solution to the integral in Equation (4.19). This is commonly

done by using random or systematic draws from the prior distribution $\pi(\beta)$ (Becerra & Goos, 2021, 2023; Kessels et al., 2009; Ruseckaite et al., 2017; Train, 2009; Yu et al., 2010). Halton draws are a practical way to obtain systematic draws from the prior distribution because they reduce the variance of the approximation to the integral and provide a good coverage of the entire domain of the prior distribution (Train, 2009; Yu et al., 2010). Denoting the number of draws by R and each individual draw by $\beta^{(i)}$, the approximation for the criterion in Equation (4.22) is

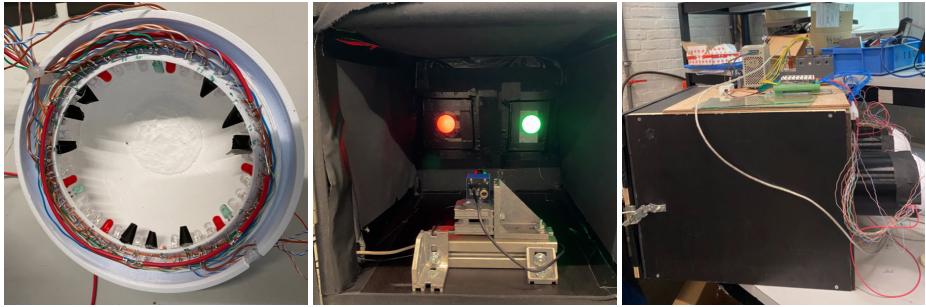
$$\mathcal{I}_B \approx \frac{1}{R} \sum_{i=1}^R \text{tr} \left[\mathbf{I}_{nc}^{-1}(\mathbf{X}_{nc}, \beta^{(i)}) \mathbf{W}_{nc} \right]. \quad (4.23)$$

It is common to use $R = 128$ Halton draws from a multivariate normal prior distribution, as mentioned in Chapters 2 and 3, as well as in Ruseckaite et al. (2017).

To compute I-optimal designs, we used a coordinate-exchange algorithm (Goos & Jones, 2011; Kessels et al., 2009; Meyer & Nachtsheim, 1995). This coordinate-exchange algorithm starts from a random initial design, and begins by optimizing the first ingredient proportion of the first alternative within the first choice set, followed by the second ingredient proportion of the first alternative within the first choice set, and so on, until all q ingredient proportions have been optimized. Then, it continues with the amount variable. The algorithm then repeats this process for each alternative in each choice set in the design. The whole process is repeated until the design can no longer be improved or until a maximum number of iterations has been reached. Details about the implementation can be found in Ruseckaite et al. (2017) and in Chapters 2 and 3.

4.4 Experimental setup

We carried out ten experiments in a black wooden box of size $40 \times 40 \times 45$ cm with two plastic domes located in the back, allowing us to show two different light mixtures at different intensities to the flies inside the box. We installed LED light bulbs of seven different colors around the edge of each dome, all pointing towards the center. Photos of the experimental setup can be seen in Figure 4.1. The colors and wavelengths of the LEDs were red (660 nm), orange (617 nm), yellow (590 nm), green (530 nm), blue (465 nm), purple (405 nm), and ultraviolet (385 nm). Additionally, the intensity of each LED could be adjusted, which we modeled as an amount variable that could take values between -1 and 1 . For each experiment, we used a batch of (approximately) eighty fruit flies. In each of the experiments, the fruit flies had to make several choices. They were presented choice sets of size two, meaning that the LED lights in the two domes were turned on with two different mixtures of colors at different intensities. In each choice set, the light was left on for twenty minutes, followed by a five minute pause with no lights on. A camera was set up to take pictures of the two domes every ten seconds, resulting in around 120 images per choice set. After each experiment was finished, we used a software that automatically counted the number of flies on each dome in each image.



(a) One of the two domes (b) Front view of the box (c) Side view of the box

Figure 4.1: Photos of the experimental setup. The first subfigure shows one of the two plastic domes with LEDs, whereas the second and third subfigures show the front and side view of the black wooden box in which the flies were allowed to fly freely.

For the first experiment, we created a Bayesian I-optimal design assuming a zero β parameter with high uncertainty for the main effects of the colors and for the interactions between the colors with the total intensity, and with low uncertainty for the pairwise interactions among different colors and for the quadratic effect of the total intensity. We did this because the interactions among the colors were not of primary interest at this initial phase. Like Ruseckaite et al. (2017) and Becerra and Goos (2021, 2023), we used $R = 128$ Halton draws from a multivariate normal prior distribution for approximating the Bayesian I-optimality criterion. The high uncertainty was conveyed through a standard deviation of 10, whereas the low uncertainty was conveyed through a standard deviation of 0.01. After this first experiment was performed, the no-choice multinomial logit model was fitted to the data. Particularly, we took the maximum number of flies in all the pictures for each alternative in each choice set, and fitted the model to this reduced data set. The model gave us an estimated mean and variance-covariance matrix, which in turn we used for the normal prior distribution of the next experiment. After the second experiment was performed, we repeated the same process using all the data we had up until that moment, and so on.

The number of choice sets for each experiment was 59, 60, 60, 60, 29, 60, 60, 72, 80, and 88. In general, we aimed for 60 choice sets per experiment. In the first and fifth experiment, some choice sets are missing because of technical problems with the set-up of the camera. For the eighth, ninth and tenth experiments, we fitted the mixed logit model outlined in Section 4.2.4 with the data that we had collected up to that stage in the experimentation process, and we manually chose 12, 20, and 28 additional choice sets involving the color combinations that had the highest predicted utilities.

4.5 Results

4.5.1 Bayesian estimation

We fitted the mixed logit model described in Section 4.2.4. Since we were working with 7 colors (the mixture variables) and one intensity variable (the amount variable), we modeled the perceived utility of alternative j in choice set s in experiment m as

$$U_{jsm} = \begin{cases} \sum_{i=1}^6 \gamma_{im}^{0*} x_{ijsm} + \sum_{i=1}^6 \sum_{k=i+1}^7 \gamma_{ikm}^0 x_{ijsm} x_{kjsm} \\ + \sum_{i=1}^7 \gamma_{im} x_{ijsm} z_{jsm} + \alpha_m z_{jsm}^2 + \varepsilon_{jsm}, & \text{for } j \in \{1, 2\}, \\ \lambda_m + \varepsilon_{jsm}, & \text{for the no-choice option,} \end{cases} \quad (4.24)$$

with λ_m the parameter for the no-choice option. Hence, the parameter vector for each experiment m is

$$\beta_m = (\gamma_{1m}^{0*}, \gamma_{2m}^{0*}, \dots, \gamma_{6m}^{0*}, \gamma_{12m}^0, \dots, \gamma_{67m}^0, \gamma_{1m}, \gamma_{2m}, \dots, \gamma_{7m}, \alpha_m, \lambda_m)^T.$$

This vector has 36 elements, including the no-choice parameter λ_m .

For the fitting process, we used the Stan programming language (Stan Development Team, 2022) through the `rstan` package (Stan Development Team, 2023) in R (R Core Team, 2017). In the data preprocessing, we discarded the first and last 5 minutes of each choice set to filter out the images with the flies recently exposed to each alternative, and then we randomly selected 30 images from each choice set to facilitate the computation process. For the priors in our model, we used a value of $\eta^* = 5$, favoring less correlation in the Ω matrix, and each element of the τ vector was set to have a mean $\mu^* = 1$ and a standard deviation $\sigma = 0.5$. The β^* vector had the values presented in the second column of Table 4.1, and for the Σ^* matrix we assumed a diagonal matrix with the values presented in the third column of Table 4.1. The values in Table 4.1 were obtained from a multinomial logit model fitted to a similar experiment performed with another species of fruit fly, the *Drosophila melanogaster*, also known as the common fruit fly, with the reasoning that their preferences would be somewhat similar to those of the *Drosophila suzukii*. We also tested a less informative prior and obtained similar estimates, but the computation times were longer and the effective sample sizes of the posterior distributions were lower, meaning that the autocorrelation of the samples of the posterior distributions was higher.

As previously mentioned, we performed ten experiments in total. The first nine experiments were used to fit the mixed logit model; and the tenth experiment was used as a confirmatory experiment to validate the model and the predictions obtained. However, before the ninth experiment was performed, we used the data from the first eight experiments to fit the mixed logit model using the priors described. This implementation gave us posterior samples of all parameters of interest. In particular, we had samples from the posterior distributions of β and Σ . With these, we obtained posterior predictive samples of the utilities of a new experiment using a grid of colors in the experimental region. By iteratively finding

parameter	term	mean	SD
γ_1^{0*}	R	-1.919	0.560
γ_2^{0*}	O	-1.612	0.563
γ_3^{0*}	Y	-2.188	0.565
γ_4^{0*}	G	-2.565	0.547
γ_5^{0*}	B	-2.652	0.536
γ_6^{0*}	P	-0.715	0.519
γ_{12}^0	$R \times O$	-0.097	0.879
γ_{13}^0	$R \times Y$	0.320	0.934
γ_{14}^0	$R \times G$	2.262	0.699
γ_{15}^0	$R \times B$	3.363	0.794
γ_{16}^0	$R \times P$	1.589	0.716
γ_{17}^0	$R \times UV$	5.848	0.651
γ_{23}^0	$O \times Y$	3.906	0.914
γ_{24}^0	$O \times G$	-2.951	1.161
γ_{25}^0	$O \times B$	0.694	0.774
γ_{26}^0	$O \times P$	0.482	0.690
γ_{27}^0	$O \times UV$	5.426	0.717
γ_{34}^0	$Y \times G$	1.266	0.739
γ_{35}^0	$Y \times B$	2.883	0.781
γ_{36}^0	$Y \times P$	2.477	0.883
γ_{37}^0	$Y \times UV$	5.284	0.756
γ_{45}^0	$G \times B$	2.494	0.961
γ_{46}^0	$G \times P$	-0.761	0.752
γ_{47}^0	$G \times UV$	2.938	0.666
γ_{56}^0	$B \times P$	-1.110	0.635
γ_{57}^0	$B \times UV$	5.217	0.487
γ_{67}^0	$P \times UV$	-0.146	0.475
γ_1	$R \times \text{intensity}$	-0.287	0.325
γ_2	$O \times \text{intensity}$	-0.705	0.487
γ_3	$Y \times \text{intensity}$	0.475	0.273
γ_4	$G \times \text{intensity}$	0.286	0.256
γ_5	$B \times \text{intensity}$	0.003	0.134
γ_6	$P \times \text{intensity}$	0.278	0.199
γ_7	$UV \times \text{intensity}$	0.175	0.217
α	intensity^2	-0.032	0.097
λ	no choice	2.213	0.337

Table 4.1: Prior mean and standard deviation of parameter vector β in the Bayesian analysis.

the combinations of colors and intensity that had the highest predicted utilities, we selected the top 5 colors, as well as the color that had the lowest predicted utility.

We then used these six colors to create one last experimental design to measure the preference of the flies for these colors. The composition of these six colors is shown in Table 4.2, and they are displayed visually in Figure 4.2. In the table

and the figure, the colors are ordered in decreasing order of expected predicted utility, meaning that, according to the model, the first color was the one with the highest utility, followed by color 2, and so on. As can be seen both in Figure 4.2 and in Table 4.2, the first three colors are very similar to each other, and they were expected to yield a similar utility. Colors 4 and 5 are also similar to each other, and, while they were expected to be less attractive than colors 1, 2, and 3, they were still expected to be attractive. Color 6 is the worst possible color to show to the flies, according to the model. The experimental design for the ninth experiment included all 15 color combinations of these six manually chosen colors. However, due to some setup problems, we did not get information for the color combinations of the pairs of colors 1 and 2, 1 and 3, 2 and 3, and 4 and 5. Therefore, information concerning these pairs was obtained in the confirmatory experiment. In the end, this is a good thing, because the confirmatory experiment was performed with data that was not used for the data fitting process, and it also includes some unseen choice sets on which to predict.

Color	Red	Orange	Yellow	Green	Blue	Purple	UV	Intensity
1	0.048	0	0	0.333	0	0	0.619	0.66
2	0.056	0	0	0.333	0	0	0.611	0.60
3	0.056	0	0	0.333	0	0	0.611	0.80
4	0.225	0.003	0	0	0	0.238	0.534	0.57
5	0.238	0	0	0	0	0.238	0.524	0.55
6	1.000	0	0	0	0	0	0	-1.00

Table 4.2: The six colors obtained from the optimization process using the model, ordered in decreasing expected utility

Once the nine experiments had been performed, we used the data from those nine experiments to fit the model using the same prior distributions as before, also filtering the first and last 5 minutes of each choice set and randomly selecting 30 images from each choice set. According to our mixed logit model, these 30 images were treated as conditionally independent from each other, the conditioning being on each parameter β_m . In the end, we had 540 choice sets, each with 30 images and 3 alternatives: left, right, and no-choice. We used 800 iterations and 4 independent chains of the MCMC algorithm implemented in Stan, with the first 300 iterations as warmup. Stan did not raise any issues or warnings about the convergence of the chains. Additionally, the convergence and mixing of the chains was inspected using the Gelman-Rubin convergence statistic (Gelman & Rubin, 1992). All values were very close to 1, indicating good mixing and convergence. Moreover, the effective sample size of the MCMC samples was high, indicating low autocorrelation within each chain (Stan Development Team, 2022).

The posterior 80% and 95% probability intervals of parameter vector β are shown in Figure 4.3. Notice in Figure 4.3 that the sizes of the effects involving intensity are much smaller than those involving the colors. That is due to the fact that the crossproducts involving only colors are much smaller than the crossproducts involving the intensity. Hence, the size of the effect should not be taken at face value. In mixture-amount experiments with a rescaled amount variable, the first-order effects are interpreted as the expected response for a product consisting

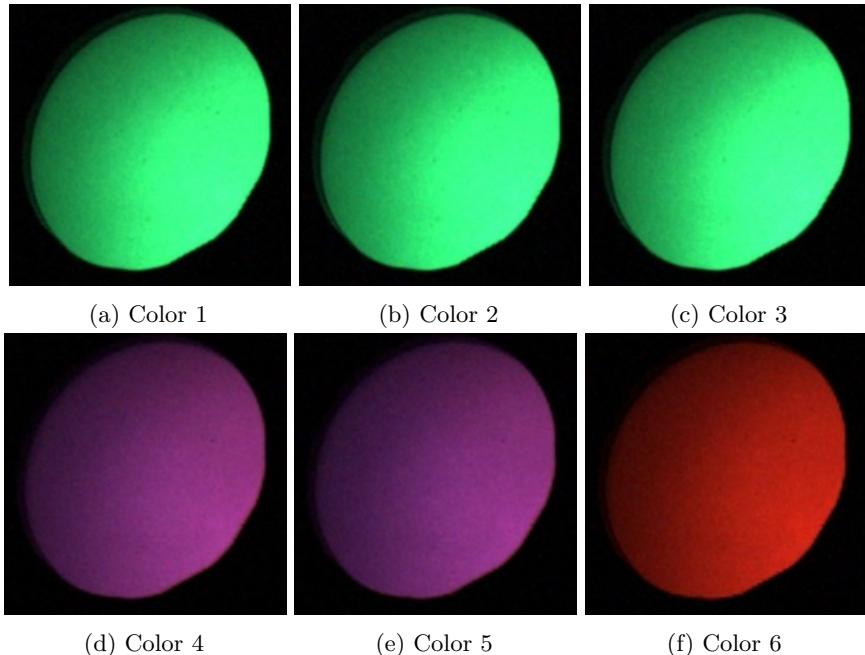


Figure 4.2: The six colors obtained from the optimization process using the model estimated with data from the first eight experiments, ordered in decreasing order of predicted expected utility. The composition of the six colors is shown in Table 4.2.

only of one given ingredient when the intensity is at its middle level (i.e., equal to zero). The interaction effects represent antagonism if their value is negative and synergism if it is positive. However, as we are working with utility and we removed one of the main-effect terms, all these effects have to be interpreted as relative to the omitted ingredient, in this case UV. So, the base level of utility corresponds to an alternative that involves 100% UV. Since all posterior intervals in the left part of Figure 4.3 are negative, using 100% of any of the other colors yields a decrease in utility, with red yielding the highest decrease. On the other hand, most of the interactions are positive and thus work synergistically to increase the utility in comparison to UV alone, the exceptions being red with yellow and yellow with blue. As for the intensity, its interactions with the colors are mostly positive, meaning that there is a positive linear effect of increasing the total intensity on the blending properties of the colors; but the quadratic effect is negative, meaning that there is a negative quadratic effect of increasing the total intensity on the blending properties of the colors, and gives a downward opening parabola, as shown in Figure 4.4. This figure shows the predicted utility for the 7 pure colors as well as for the best predicted color (color 1 in Table 4.2 and Figure 4.2). Each line represents each of these colors, and they all show an inverse U-shape, with 100% red being the least attractive color, and color 1 being the most attractive one.

Figure 4.5 shows the 95% probability intervals of parameter vector β , as well as

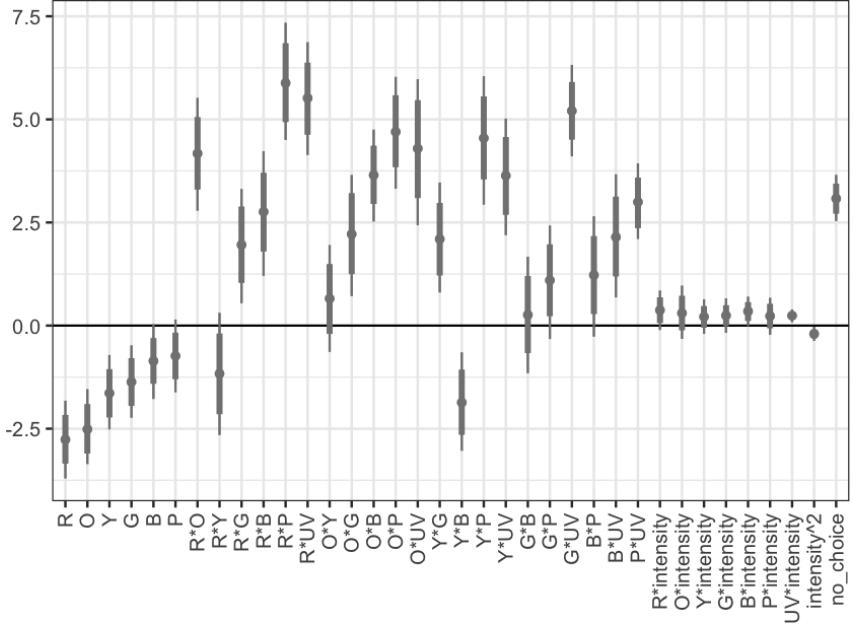


Figure 4.3: Posterior probability intervals of parameter vector β . The dots represent the median, the thick lines the 80% interval, and the thin lines 95% interval.

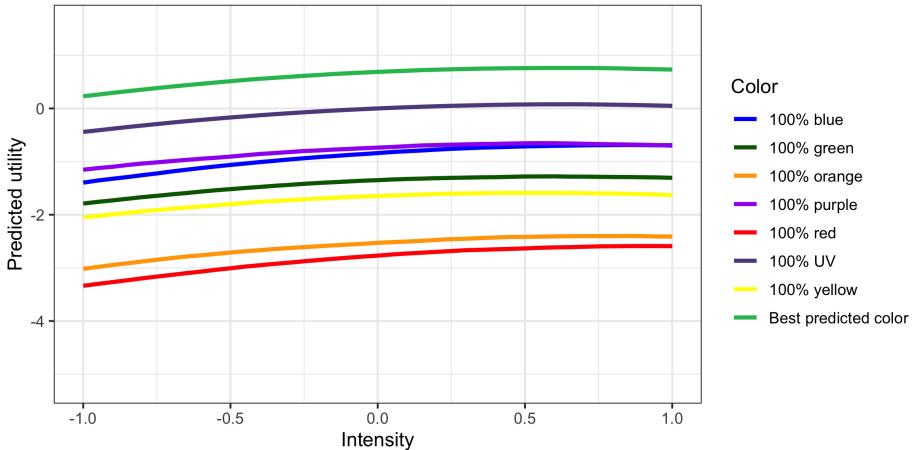


Figure 4.4: Predicted utility for light with 100% of each color and varying intensity. Each line represents a color, with the best predicted color being color 1 in Table 4.2 and Figure 4.2. All lines are downward-opening parabolas.

the intervals of the parameter vectors β_m , for m from 1 to 9. The red intervals correspond to β , whereas the gray ones correspond to each β_1, \dots, β_9 . We can see some outliers with lower values and wider intervals than the rest, particularly for the interaction of orange with yellow, orange with green, orange with UV, yellow

with UV, and green with blue. These intervals correspond to experiments 1 and 5, both of which, as previously mentioned, had some issues. Experiment 1 was more focused on first-order effects, so it is not surprising that there are some outliers in these estimates. Additionally, this experiment had a particular set of inactive flies. Experiment 5 had 31 missing choice sets due to some technical problems. Therefore, it was less informative, which led to lower point estimates and wider intervals.

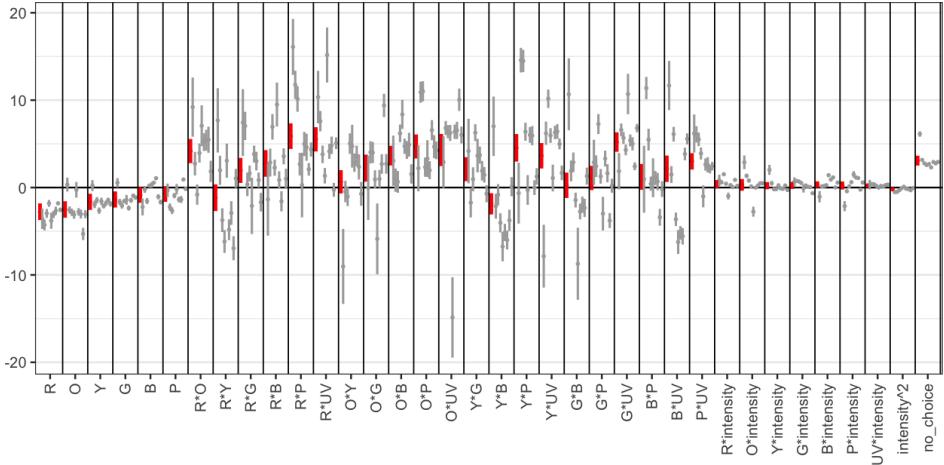


Figure 4.5: Posterior 95% probability intervals of parameter vector β (red) and vectors β_m (gray).

4.5.2 Model validation

To check the fit of the model, we performed posterior predictive checks of the data, which work by replicating data under the model and comparing it to the observed data. To do this, we simulated values from the joint posterior predictive distribution of replicated data to compare these samples with the observed data. Posterior predictive checks combine uncertainty about parameters with uncertainty about outcomes, the former by the posterior distribution and the latter by the likelihood function (McElreath, 2018). The idea behind this is that if a model is a good fit, then we should be able to use it to generate data that looks like the observed data (Gabry et al., 2019). If we denote by \mathbf{y}^{rep} a vector of replicated data, then the technical definition of its distribution would be

$$p(\mathbf{y}^{\text{rep}} | \mathbf{X}, \mathbf{y}) = \int p(\mathbf{y}^{\text{rep}} | \boldsymbol{\theta}, \mathbf{X}) p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) d\boldsymbol{\theta}. \quad (4.25)$$

In practice, creating the replicated data is quite simple. From each of the samples obtained from the posterior distribution $p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})$, which was defined in Equations (4.17) and (4.18), we can compute a replication of the response vector \mathbf{y} . In particular, for each sample r from β_m , which we denote by $\beta_m^{(r)}$, we compute the probability of choosing alternative j in choice set s in experiment m using

Equation (4.9). We denote this probability by $p_{jsm}^{(r)}$. Then, we simulate how many flies choose each alternative and how many choose the no-choice option by sampling from a multinomial distribution with probability vector $(p_{1sm}^{(r)}, p_{2sm}^{(r)}, p_{3sm}^{(r)})^T$ and size 80. This simulation gives us a replicated response vector $(y_{1sm}^{(r)}, y_{2sm}^{(r)}, y_{3sm}^{(r)})^T$ for choice set s and experiment m . If we concatenate all such replicated response vectors for all choice sets in all experiments, we obtain a single replicated response vector, $\mathbf{y}^{(r)}$, for the whole data set that was used to fit the model. Each vector $\mathbf{y}^{(r)}$ is a sample from the distribution defined in Equation (4.25).

Using the replicated data, we can check the fit of the data analytically and visually. For each experiment m , first we computed the number of flies choosing each alternative in each experiment, and compared it visually with the observed values. This is shown in Figure 4.6, which was created using the `bayesplot` package (Gabry & Mahr, 2022). In each subfigure, the bars represent the histogram of the numbers of flies that were observed choosing each alternative in each experiment, whereas the points and lines on top of them show the median and the 90% intervals of the posterior predictive distribution for the number of flies. In eight of the nine experiments, the model reproduces the fitted data well, confirming a good model fit. The exception is the first experiment. In the left alternative of experiment 1, most of the observed responses were 0 and 1, meaning that, most of the time, no fly or only one fly was attracted to the alternative on the left. Our model produces something similar, with most of the simulated data sets having $y_{1sm}^{(r)}$ and $y_{2sm}^{(r)}$ equal to 0 or 1. Something similar happens with the alternative on the right for experiment 1, except that in this case there were not as many zeros as in the alternative on the left. Our model actually fails to capture this because it overestimates the number of zeros and underestimates the number of ones. This could either mean that there is an effect that we are not including in the model that would capture this behavior, such as the fact that this first experiment was performed in winter, where flies might be less active because of the temperatures and decreased sunlight, whereas the rest of the experiments were performed in late spring and in summer. Or it could be random noise.

4.5.3 Model predictions

As previously mentioned, an additional confirmatory experiment was performed. The data from this confirmatory experiment were not used in the model analysis, but rather it was used to confirm whether the predictions from the model about the preferences of the flies were correct. Since it was not used in the data modeling, this last experiment's dataset can be seen as a holdout set to test the model's predictive performance. This experiment had 88 choice sets, of which 30 were particularly interesting: all the combinations of the 6 selected colors in Table 4.2 and Figure 4.2, with each color on the left once and on the right once. These choice sets were randomly interspersed with the other choice sets in the experiment. As with the previous experiments, we obtained around 120 images per choice set.

Using our model, we obtained posterior predictive intervals for the number of flies that each alternative in each choice set would attract. These intervals were

obtained using the posterior predictive distribution. In general, for a new point of interest \mathbf{a}_{new} , the posterior predictive distribution of the response \mathbf{y}_{new} is

$$p(\mathbf{y}_{\text{new}} | \mathbf{X}, \mathbf{y}, \mathbf{a}_{\text{new}}) = \int p(\mathbf{y}_{\text{new}} | \boldsymbol{\theta}, \mathbf{a}_{\text{new}}) p(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) d\boldsymbol{\theta}. \quad (4.26)$$

Notice that the definition in Equation (4.26) is similar to that of the replicated responses for the posterior predictive checks in Equation (4.25), but it conditions on the new point of interest \mathbf{a}_{new} . The approach to compute the intervals is similar to the posterior predictive checks, but here we are predicting for a new experiment, not for an observed one. This means that, in general, the new intervals will be wider, and that extra steps are needed to handle the draws from the posterior distributions of the parameters. We do not use draws from the posterior distribution of $\boldsymbol{\beta}_m$, but rather from the posterior of $\boldsymbol{\beta}$. This is because this is a prediction for a new experiment. To do this, for each draw r from $\boldsymbol{\beta}$, which we denote by $\boldsymbol{\beta}^{(r)}$, and for each draw from Σ , which we denote by $\Sigma^{(r)}$, we obtain a draw for the new experiment's parameter, which we denote by $\boldsymbol{\beta}_{\text{new}}^{(r)}$. This draw, denoted $\boldsymbol{\beta}_{\text{new}}^{(r)}$, is obtained by sampling from a multivariate normal with mean $\boldsymbol{\beta}^{(r)}$ and variance-covariance matrix $\Sigma^{(r)}$, as per Equation (4.10). The draw $\boldsymbol{\beta}_{\text{new}}^{(r)}$ is then used to compute the probability of choosing alternative j in choice set s in the new experiment using Equation (4.9), which is then used in the multinomial distribution to obtain a response, as in the posterior predictive checks.

Using this procedure, we computed the 80%, 60%, 50%, 40%, and 10% posterior predictive intervals, as well as the median, which we then compared with the actual number of flies that were attracted to each alternative in each choice set. Figure 4.7 shows the posterior predictive intervals, as well as the number of flies in each image. Lighter tones of gray represent larger intervals, while darker tones represent smaller intervals. The darkest gray corresponds to the posterior median. Each of the points shows the number of flies in an image. Each subplot represents one of the 30 choice sets with their left and right alternatives, the L denoting *left* and the R denoting *right*. The number next to the colon refers to the color shown in that alternative (see Table 4.2 for the composition of each color). The horizontal black lines show the observed median number of flies that were attracted to each alternative in each choice set. The observed median is shown at the top.

For example, the first choice set is comprised of colors 1 and 2. The model gives almost the same predictions for the number of flies picking each color, with the 80% prediction interval being relatively large, going from 0 to 30. The 60% interval goes from 1 to 21. The 50% interval ranges from 2 to 18 for color 1 and from 2 to 17 for color 2. The 40% interval goes from 2 to 15 for color 1 and from 2 to 14 for color 2. Finally, the 10% interval ranges from 5 to 8 for both colors, with the dark gray line being the predicted median, which is 7 for both colors. The actual number of flies that were attracted to the two alternatives in this choice set ranges from 0 to 6, with color 1 (left) having a median of 3 and color 2 (right) having a median of 2, shown with a black horizontal line as well as the text at the top. In this case, the actual number of flies was lower than what the model predicted. However, the number of flies attracted to both colors was roughly the same, just as the model predicted. A similar situation occurs for choice sets 2, 3,

and 4. In choice set 5, our predictions are more in line with the actual number of flies attracted to the alternatives.

Figure 4.7 also shows that in choice sets 9 and 10, for example, the prediction for color 6, which is the most unattractive, is correct in the sense that the model predicted that almost no flies would be attracted to that alternative. Indeed, the median number of flies is zero for color 6 in choice sets 9 and 10. However, we can see that the number of flies for color 1, which is the most attractive according to the model, varies considerably from one choice set to the other, with color 1 attracting a median of 3 flies with relatively little variation in choice set 9; and a median of 13 flies in choice set 10, with the observed number of flies ranging from 0 to 22. This difference in particular seems to be due to the fact that choice set 9 was presented relatively early in the experiment, and the flies did not seem to be very active at that time, while choice set 10 was presented much later in the experiment, when the flies seemed to be more active. In any case, the large number of active flies in choice set 10 is an outlier, it being the choice set with the highest number of flies that chose a color in the whole experiment.

Something similar happens in choice sets 5 and 6: the predictions for choice set 5 are more accurate than for choice set 6, and it turns out that, in choice set 6, the number of attracted flies in general is lower than in choice set 5, despite having the same colors. This could be due to the time of the day at which each of the choice sets were presented, the temperature of the room, or just the randomness in the behavior of the flies. Rice et al. (2016) noticed different preferences of *Drosophila suzukii* in two different seasons. This suggests that there might be variables that we could include in our model to improve its predictive performance. However, once again, the model does correctly predict the preference of the flies for each of the alternatives, which is that color 1 is more attractive than color 4. In fact, this correct prediction happens in all but 4 of the choice sets in Figure 4.7. In choice sets 1, 12, 13, and 15, our prediction of which of the two alternatives would attract more flies is wrong, when comparing the median number of flies in each alternative. For example, in choice set 1, the model predicts that both alternatives would attract the same number of flies, but in reality the median number of flies attracted to left alternative is higher than the median number of flies attracted to the right. Interestingly, in choice set 2, which involves the same colors, the prediction is correct. The same happens with choice sets 12, 13 and 15: the choice set with the same colors in reverse order is correctly predicted by the model.

The general conclusion from Figure 4.7 is that the predictions of which of the two colors presented is more attractive are generally correct. The model correctly predicts which alternative is more attractive in 26 of the 30 choice sets shown in Figure 4.7. In addition, the prediction is also correct in 47 of the 58 remaining choice sets in the confirmatory experiment, but not shown here. The model's prediction intervals for the number of flies are relatively wide and could perhaps be reduced by including some other variables in the model. However, in this first series of experiments, the goal was to find the colors that are the most attractive to the flies, and not necessarily accurately predict the number of flies, or model the exact behavior of the flies. With this goal in mind, we can confidently say that our model works well.

Additionally, it should be noted that the predictions we performed for this confirmatory set were done on raw data. The data that we fed to the model were preprocessed by filtering the first and last images from each choice set, and also randomly sampling 30 images for each choice set. Also, due to the nature of the data, the raw data has some time autocorrelation, which was not included in the model. In fact, the images were treated as conditionally independent from each other, the conditioning being on each parameter β_m . Therefore, it is remarkable that the model predicts the new data so well.

4.6 Discussion

In this chapter, we described the design, set-up, and data analysis of several mixture-amount choice experiments that were intended to gain insight into the color preferences of the *Drosophila suzukii* fruit fly. This was done with the objective of developing traps that are of an attractive color for the flies. This way, the presence of this pest can be monitored in orchards and other places where fruit is grown. We modeled the color preferences of the flies through a mixed logit model which helped us identify the most attractive combinations of colors, and we confirmed their attractiveness through a confirmatory experiment. We found that *Drosophila suzukii* are attracted to UV light, but a combination of UV light and other colors seems to be more attractive to the flies than pure UV light. In fact, according to our model, the most attractive color is a mixture of 4.8% red, 33.3% green, and 61.9% UV; with an intensity of 0.66 on a scale from -1 to 1 with our particular light setup.

Our model was mostly able to correctly discern which of the two alternatives shown to the flies was the most attractive in the holdout set, although the prediction about the exact number of flies that were attracted to each alternative was not very precise. However, we argue that the objective of this work was to have a model that would be able to identify attractive color combinations. As Gelman et al. (2013) state, instead of asking ourselves if our model is true or false, we should ask ourselves the more relevant question of whether our model's deficiencies have a noticeable effect on the substantive inferences. We can confidently say that, for the inferences of our consideration, our model's deficiencies did not have a noticeable effect.

On the other hand, correctly predicting the number of flies attracted to each color might be of interest for the development of traps. A subsequent experiment might want to model the preferences of flies and correctly predict the number of flies that would be attracted in order to find the color that attracts the highest number of flies. We believe that the reason our model lacks power in that particular sense is because there are variables that could be included in the model that would aid in the prediction. For example, the temperature and humidity might affect the flies' activity. Hamby et al. (2013) performed experiments on *Drosophila suzukii* simulating summer and winter scenarios in California, by modifying the temperature, humidity, and amount of light. They found that the flies' behavior during the simulated summer had a bimodal pattern of activity, with an afternoon

“nap” during the period with the highest light intensity and warmest temperature; but that during the simulated winter the behavior was unimodal with an activity peak in the middle of the day with the highest light intensity and temperature. Jaffe and Guédot (2019) also noted different peaks of activity at different times of the day and behavior that changed with temperature and humidity for flies in raspberry patches. Including these type of variables and studying their interaction with the colors might be of interest and a future direction of work. Additionally, if the same setup as ours is used, perhaps modeling the time autocorrelation might help improve the predictions.

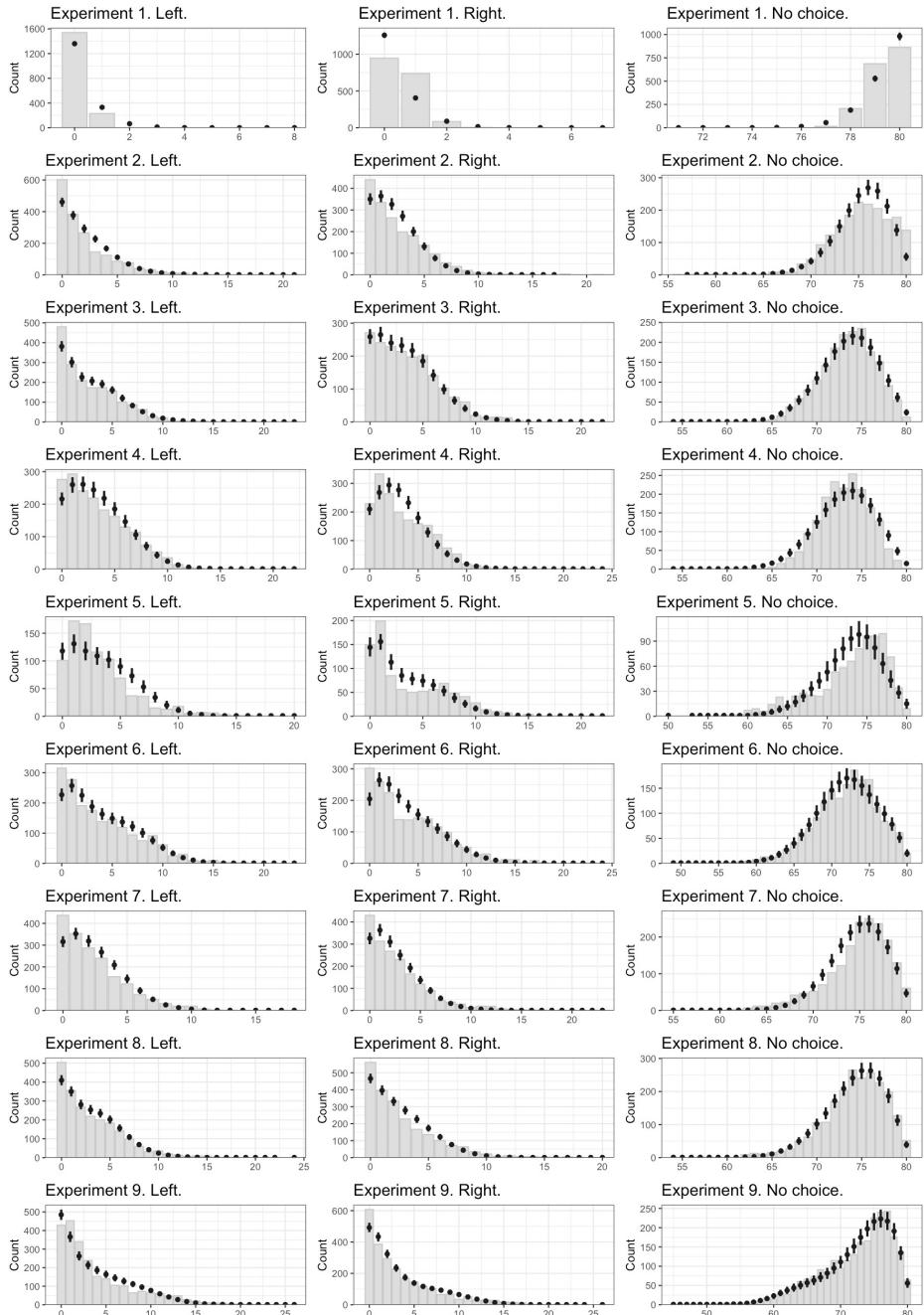


Figure 4.6: Posterior predictive checks of counts. In each subfigure, the bars represent the histogram of the numbers of flies that were observed choosing each alternative in each experiment, whereas the points and lines on top of them show the median and the 90% intervals of the posterior predictive distribution for the number of flies.

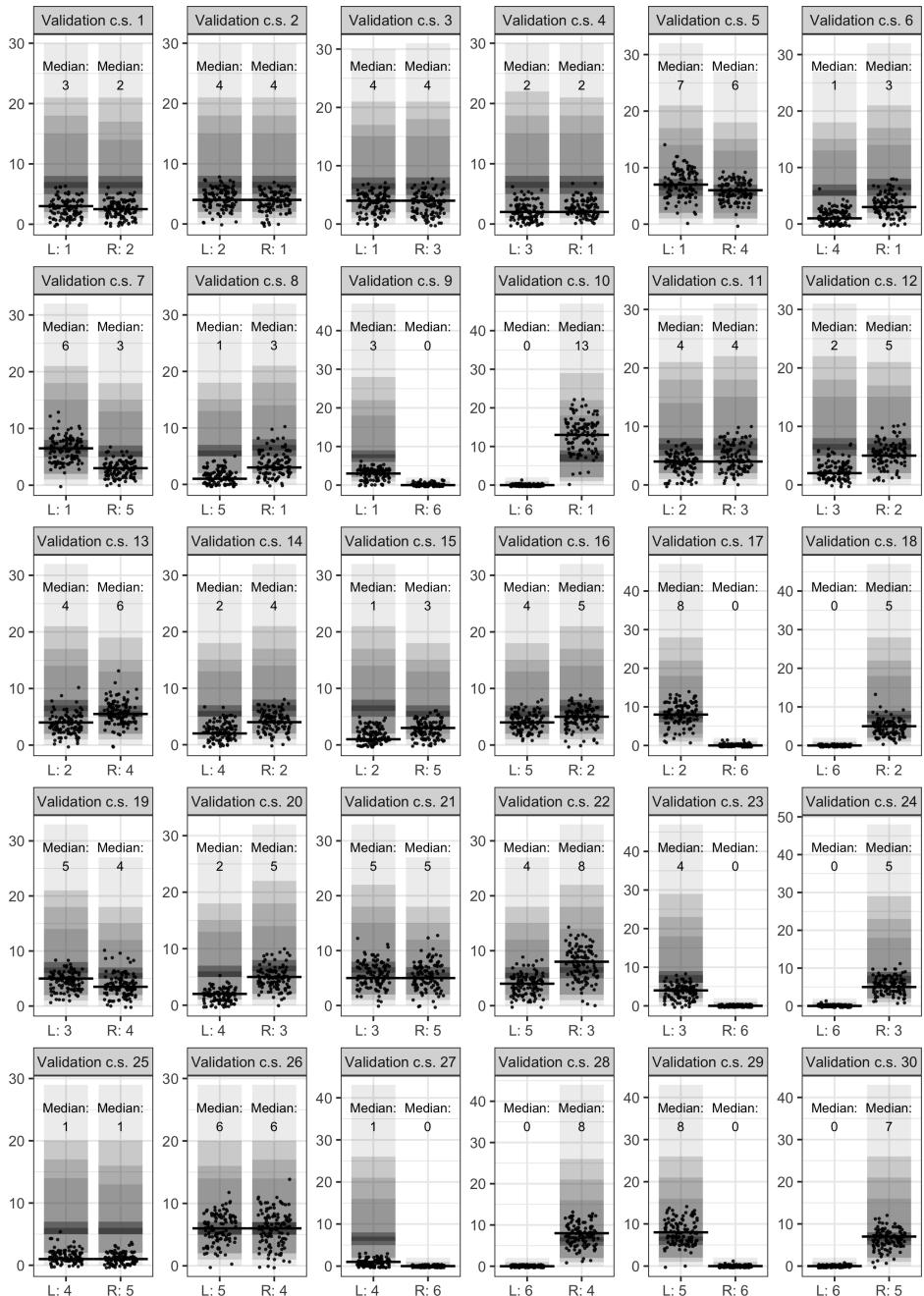


Figure 4.7: Posterior predictive intervals for 30 of the 88 choice sets in the confirmatory experiment. Lighter tones of gray represent larger intervals, while darker tones represent smaller intervals. The darkest gray corresponds to the posterior median

- Each of the points shows the number of flies in an image. The number next to the colon refers to the color shown in that alternative (see Table 4.2). The horizontal black lines show the observed median number of flies that were attracted to each alternative in each choice set.

Chapter 5

Conclusions and future work

5.1 Conclusions

In this thesis, we showed how to combine mixture-process variable experiments and discrete choice experiments to quantify and model preferences for products, articles, or other items that can be viewed as mixtures of ingredients. We introduced a computationally efficient definition for Bayesian I-optimal designs for choice experiments involving mixtures of ingredients with and without process variables. We embedded the new criterion in a coordinate-exchange algorithm for constructing Bayesian D- and I-optimal choice designs and demonstrated through different examples that the I-optimal designs perform substantially better than their D-optimal counterparts in terms of the variance of the predicted utility. We also applied the concepts of optimal design and analysis of mixture-process variable choice experiments in a series of experiments involving fruit flies. We designed and carried out several I-optimal choice experiments with fruit flies to model their color preferences and identified the most attractive combinations of colors to the flies.

In Chapter 2, we introduced a computationally efficient definition for Bayesian I-optimal designs for choice experiments involving mixtures of ingredients. We embedded the new criterion in a coordinate-exchange algorithm for constructing Bayesian D- and I-optimal designs and demonstrated that the I-optimal designs perform substantially better than their D-optimal counterparts in terms of the variance of the predicted utility through two examples.

In Chapter 3, extended the work from Chapter 2. This extension was inspired by the fact that the preference for a mixture may depend on characteristics other than its composition. For example, the most preferred bread might not only depend on the proportions of the various ingredients, but also on the baking time and the baking temperature; or the ideal cocktail may also depend on the temperature at which it is served. To cope with this kind of complication, we extended our model to deal with these additional characteristics, typically called *process variables*.

(Goos & Jones, 2011). We introduced the theory for choice experiments involving mixtures and process variables, and embedded the Bayesian D- and I-optimality criteria in a coordinate-exchange algorithm for constructing optimal designs for this type of choice experiments.

In Chapter 4, we described the design, set-up, and data analysis of several mixture-amount choice experiments that were intended to gain insight into the color preferences of the species of fruit fly *Drosophila suzukii*. This was done with the objective of developing traps of an attractive color for the flies. This way, the presence of this pest can be monitored in orchards and other places where fruit is grown. We modeled the color preferences of the flies through a mixed logit model which helped us identify the most attractive combinations of colors, and confirm their attractiveness through a confirmatory experiment. We found that *Drosophila suzukii* are attracted to UV light, but a combination of UV light and other colors seems to be more attractive to the flies than pure UV light. In fact, according to our model, the most attractive color is a mixture of 4.8% red, 33.3% green, and 61.9% UV; with an intensity of 0.66 on a scale from –1 to 1 with our particular light setup.

5.2 Future work

We identified several possible extensions of the work presented in this thesis. The first thing would be to extend the work to other classes of models for data from mixture experiments than the Scheffé models, for example Becker models or Cox's mixture polynomial models (Cornell, 2002).

Second, in our designs we focused on the multinomial logit model, which assumes that there is homogeneity in the preferences of the respondents. However, as demonstrated by Courcoux and Sémenou (1997b) and Goos and Hamidouche (2019), this might be an unrealistic assumption. Therefore, it would make sense to extend the models and algorithms presented here to other types of models that take into account the possible presence of consumer heterogeneity, such as the mixed logit model presented in Chapter 4. One way to do this could be by doing individually adapted sequential Bayesian designs. Yu et al. (2011) presented an algorithm to create this type of designs by presenting each individual respondent their own personalized choice sets based on their previous answers. The algorithm assumes that respondents choose according to a multinomial logit model where each respondent has different preferences. Danthurebandara et al. (2011) followed this same approach but for an application focused on estimating the willingness-to-pay of consumers. Crabbe et al. (2014) studied this further by comparing the computation time of different optimality criteria. The ideas and algorithms from these articles can be adapted to work with choice experiments involving mixtures and process variables.

A third opportunity for future research is inspired by a practical difficulty that arises when conducting choice experiments with mixtures either with or without process variables. When the number of distinct mixtures appearing in the Bayesian optimal designs is large and the mixtures have to be tasted, it is logically

very cumbersome to organize the experiment. For instance, organizing a choice experiment in which 40 distinct mixtures have to be tasted in perhaps 80 different choice sets is much harder to organize than a choice experiment in which only 20 distinct mixtures have to be tasted in 40 different choice sets. While the former experiment may be preferable from a statistical viewpoint, it may be practically infeasible. Developing an algorithm to find Bayesian I-optimal designs with mixtures with an upper bound on the number of distinct mixtures and/or an upper bound on the number of distinct choice sets is therefore valuable from a practitioner's point of view.

Two ways to do this come to our minds. The first one is to start with the optimal design returned by our coordinate-exchange algorithm, then cluster the points such that the number of clusters is the same as the maximum number of mixtures allowed, and then run the coordinate-exchange algorithm again with a modification such that all the points within the same cluster move at the same time and continue satisfying the constraints. The second way is a very different one and involves a variable neighborhood search (VNS) approach. The VNS is a metaheuristic approach to solve complex optimization problems that has been successfully used in the optimal design of experiments (Goos et al., 2020; Syafitri et al., 2015).

A fourth topic for future research would be to modify our coordinate-exchange algorithm, so that it can also cope with experimental regions that are not a simplex. Such experimental regions arise when there are constraints on the ingredient proportions other than lower bounds for individual proportions. Methodologically speaking, this is not highly innovative, since the mixture coordinate-exchange algorithm of Piepel et al. (2005) for linear regression models is already able to deal with this complication. However, embedding this capability in our implementation of the coordinate-exchange algorithm for choice experiments with mixtures would be useful for practitioners.

A fifth area of future work is directly related to Chapter 4. If an improvement of the predictive capacities of our mixed logit model is sought, the addition of certain variables could help in the prediction. For example, the temperature and humidity might affect the flies' activity. Additionally, if the same experimental setup as ours is used, modeling the time autocorrelation might help improve the predictions.

Moreover, the combination of Scheffé models with logit type choice models can also be used to study the preferences of other animals and pests, especially those that rely on visual cues, such as the African fig fly, *Zaprionus indianus* (Cruz-Esteban, 2021). The applications of these models do not apply only to colors, but they can also be used for olfactory traps and to find the most attractive combination of ingredients, like the proportion of apple cider vinegar, rice vinegar, ethanol, yeast, sugar, flour, water, and even wine (Cha et al., 2017; Iglesias et al., 2014). Likewise, our models and experimental setup could be used to find the most attractive combination of both visual and olfactory ingredients. They could also be used to test whether the contrast between two colors attracts more flies. Finally, after more information is obtained about the flies' preferences, the experiments could be applied in fields and orchards to test whether the preferences the lab experiments generalize to the places where the flies might naturally live.

Appendix A

Appendix

A.1 Design tables of artificial sweetener experiment

Choice set	x_1	x_2	x_3
1	0.00	1.00	0.00
1	0.61	0.39	0.00
2	0.00	1.00	0.00
2	0.24	0.30	0.47
3	0.00	0.00	1.00
3	0.00	0.67	0.33
4	1.00	0.00	0.00
4	0.40	0.60	0.00
5	0.22	0.44	0.33
5	0.59	0.00	0.41
6	0.40	0.00	0.60
6	1.00	0.00	0.00
7	0.00	0.33	0.67
7	0.48	0.25	0.27

Table A.1: Bayesian D-optimal design for the artificial sweetener experiment, when $\kappa = 0.5$

Choice set	x_1	x_2	x_3
1	0.00	0.00	1.00
1	0.39	0.34	0.27
2	0.00	1.00	0.00
2	0.02	0.00	0.98
3	1.00	0.00	0.00
3	0.10	0.70	0.21
4	0.00	0.00	1.00
4	0.52	0.00	0.48
5	0.00	0.00	1.00
5	0.39	0.34	0.27
6	0.52	0.48	0.00
6	0.17	0.15	0.68
7	0.06	0.00	0.94
7	0.00	0.51	0.49

Table A.2: Bayesian I-optimal design for the artificial sweetener experiment, when $\kappa = 0.5$

Choice set	x_1	x_2	x_3
1	0.00	0.00	1.00
1	0.28	0.28	0.43
2	0.56	0.44	0.00
2	0.28	0.30	0.42
3	0.00	1.00	0.00
3	0.00	0.48	0.52
4	1.00	0.00	0.00
4	0.51	0.49	0.00
5	0.00	0.68	0.32
5	0.42	0.32	0.26
6	1.00	0.00	0.00
6	0.50	0.00	0.50
7	0.56	0.00	0.44
7	0.26	0.41	0.32

Table A.3: Bayesian D-optimal design for the artificial sweetener experiment, when $\kappa = 5$

Choice set	x_1	x_2	x_3
1	0.00	0.00	1.00
1	0.47	0.00	0.53
2	0.27	0.43	0.31
2	0.08	0.00	0.92
3	1.00	0.00	0.00
3	0.39	0.36	0.25
4	0.42	0.27	0.31
4	0.00	0.09	0.91
5	0.48	0.52	0.00
5	0.18	0.19	0.62
6	0.00	1.00	0.00
6	0.33	0.42	0.24
7	0.00	0.00	1.00
7	0.00	0.51	0.49

Table A.4: Bayesian I-optimal design for the artificial sweetener experiment, when $\kappa = 5$

Choice set	x_1	x_2	x_3
1	0.00	0.00	1.00
1	0.25	0.24	0.51
2	0.00	1.00	0.00
2	0.00	0.57	0.43
3	0.55	0.00	0.45
3	1.00	0.00	0.00
4	0.54	0.00	0.46
4	0.28	0.38	0.34
5	0.57	0.43	0.00
5	1.00	0.00	0.00
6	0.55	0.45	0.00
6	0.31	0.33	0.37
7	0.00	0.68	0.32
7	0.37	0.36	0.27

Table A.5: Bayesian D-optimal design for the artificial sweetener experiment, when $\kappa = 10$

Choice set	x_1	x_2	x_3
1	0.00	0.00	1.00
1	0.39	0.01	0.60
2	0.51	0.25	0.24
2	0.98	0.00	0.02
3	0.00	0.10	0.90
3	0.38	0.26	0.36
4	0.20	0.28	0.53
4	0.44	0.55	0.01
5	0.27	0.41	0.32
5	0.12	0.00	0.88
6	0.00	0.00	1.00
6	0.01	0.41	0.58
7	0.00	0.96	0.04
7	0.27	0.52	0.21

Table A.6: Bayesian I-optimal design for the artificial sweetener experiment, when $\kappa = 10$

Choice set	x_1	x_2	x_3
1	1.00	0.00	0.00
1	0.68	0.12	0.20
2	0.00	0.00	1.00
2	0.16	0.17	0.67
3	0.51	0.00	0.49
3	0.37	0.28	0.35
4	0.00	1.00	0.00
4	0.00	0.66	0.34
5	0.00	0.51	0.49
5	0.29	0.33	0.38
6	0.47	0.53	0.00
6	0.41	0.31	0.28
7	0.31	0.69	0.00
7	0.00	1.00	0.00

Table A.7: Bayesian D-optimal design for the artificial sweetener experiment, when $\kappa = 30$

Choice set	x_1	x_2	x_3
1	0.10	0.59	0.31
1	0.00	0.32	0.68
2	0.38	0.00	0.62
2	0.66	0.08	0.26
3	0.24	0.54	0.22
3	0.00	0.81	0.19
4	0.80	0.00	0.20
4	0.56	0.25	0.19
5	0.06	0.26	0.68
5	0.00	0.00	1.00
6	0.00	0.00	1.00
6	0.26	0.07	0.67
7	0.44	0.50	0.06
7	0.29	0.36	0.35

Table A.8: Bayesian I-optimal design for the artificial sweetener experiment, when $\kappa = 30$

A.2 Design tables of cocktail experiment

In both tables, x_1 , x_2 , and x_3 denote the pseudocomponents that range from 0 to 1, while a_1 , a_2 , and a_3 are the ingredient proportions with their original lower bounds.

Choice set	x_1	x_2	x_3	a_1	a_2	a_3
1	0.00	1.00	0.00	0.30	0.60	0.10
1	0.60	0.40	0.00	0.57	0.33	0.10
2	1.00	0.00	0.00	0.75	0.15	0.10
2	0.46	0.00	0.54	0.51	0.15	0.34
3	0.00	0.00	1.00	0.30	0.15	0.55
3	0.00	0.55	0.45	0.30	0.40	0.30
4	0.00	0.00	1.00	0.30	0.15	0.55
4	0.00	0.55	0.45	0.30	0.40	0.30
5	0.60	0.40	0.00	0.57	0.33	0.10
5	0.00	1.00	0.00	0.30	0.60	0.10
6	1.00	0.00	0.00	0.75	0.15	0.10
6	0.40	0.60	0.00	0.48	0.42	0.10
7	0.00	1.00	0.00	0.30	0.60	0.10
7	0.00	0.41	0.59	0.30	0.34	0.36
8	1.00	0.00	0.00	0.75	0.15	0.10
8	0.31	0.36	0.33	0.44	0.31	0.25
9	0.00	0.50	0.50	0.30	0.38	0.32
9	0.40	0.29	0.31	0.48	0.28	0.24
10	0.00	1.00	0.00	0.30	0.60	0.10
10	0.36	0.33	0.31	0.46	0.30	0.24
11	0.00	0.50	0.50	0.30	0.38	0.32
11	0.40	0.29	0.31	0.48	0.28	0.24
12	0.40	0.60	0.00	0.48	0.42	0.10
12	1.00	0.00	0.00	0.75	0.15	0.10
13	0.49	0.00	0.51	0.52	0.15	0.33
13	0.00	0.48	0.52	0.30	0.37	0.33
14	0.52	0.00	0.48	0.53	0.15	0.32
14	0.27	0.40	0.33	0.42	0.33	0.25
15	0.00	0.00	1.00	0.30	0.15	0.55
15	0.53	0.00	0.47	0.54	0.15	0.31
16	0.54	0.46	0.00	0.54	0.36	0.10
16	0.30	0.33	0.38	0.43	0.30	0.27

Table A.9: Bayesian D-optimal design for the cocktail experiment

Choice set	x_1	x_2	x_3	a_1	a_2	a_3
1	0.00	0.00	1.00	0.30	0.15	0.55
1	0.58	0.00	0.42	0.56	0.15	0.29
2	1.00	0.00	0.00	0.75	0.15	0.10
2	0.36	0.64	0.00	0.46	0.44	0.10
3	0.00	0.00	1.00	0.30	0.15	0.55
3	0.58	0.00	0.42	0.56	0.15	0.29
4	0.03	0.97	0.00	0.31	0.59	0.10
4	0.29	0.28	0.43	0.43	0.27	0.29
5	0.00	0.00	1.00	0.30	0.15	0.55
5	0.00	0.63	0.37	0.30	0.43	0.27
6	0.00	0.00	1.00	0.30	0.15	0.55
6	0.30	0.34	0.36	0.43	0.30	0.26
7	0.00	0.00	1.00	0.30	0.15	0.55
7	0.00	0.63	0.37	0.30	0.43	0.27
8	0.00	1.00	0.00	0.30	0.60	0.10
8	0.65	0.35	0.00	0.59	0.31	0.10
9	1.00	0.00	0.00	0.75	0.15	0.10
9	0.30	0.16	0.53	0.44	0.22	0.34
10	0.00	0.49	0.51	0.30	0.37	0.33
10	0.57	0.43	0.00	0.56	0.34	0.10
11	0.00	0.00	1.00	0.30	0.15	0.55
11	0.30	0.34	0.36	0.43	0.30	0.26
12	0.00	0.00	1.00	0.30	0.15	0.55
12	0.30	0.34	0.36	0.43	0.30	0.26
13	0.45	0.00	0.55	0.50	0.15	0.35
13	0.23	0.53	0.24	0.40	0.39	0.21
14	0.00	0.41	0.59	0.30	0.34	0.36
14	0.50	0.19	0.31	0.52	0.24	0.24
15	0.00	0.00	1.00	0.30	0.15	0.55
15	0.49	0.51	0.00	0.52	0.38	0.10
16	0.00	0.00	1.00	0.30	0.15	0.55
16	0.30	0.34	0.36	0.43	0.30	0.26

Table A.10: Bayesian I-optimal design for the cocktail experiment

Bibliography

- Agresti, A. (2012). *Categorical data analysis* (Vol. 792). John Wiley & Sons.
- Akinc, D., & Vandebroek, M. (2018). Bayesian estimation of mixed logit models: Selecting an appropriate prior for the covariance matrix. *Journal of Choice Modelling*, 29, 133–151.
- Allenby, G., Rossi, P., & McCulloch, R. (2005). Hierarchical Bayes Models: A Practitioners Guide. *SSRN Electronic Journal*. <https://ssrn.com/abstract=655541>
- Álvarez, I., Niemi, J., & Simpson, M. (2014). Bayesian inference for a covariance matrix. *arXiv preprint arXiv:1408.4050*.
- Apellado-Buenaventura, A., & Valmorida, J. S. (2021). Optimization of buffalo milk pastillas de leche using sensory acceptability criteria by Plackett-Burman and central composite designs. *International Journal of Scientific & Technology Research*, 10, 136–148.
- Asplen, M. K., Anfora, G., Biondi, A., Choi, D.-S., Chu, D., Daane, K. M., Gibert, P., Gutierrez, A. P., Hoelmer, K. A., Hutchison, W. D., et al. (2015). Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. *Journal of Pest Science*, 88, 469–494.
- Assele, S. Y., Meulders, M., & Vandebroek, M. (2023). Sample size selection for discrete choice experiments using design features. *Journal of choice modelling*, 49, 100436.
- Atkinson, A. C., & Haines, L. M. (1996). Designs for nonlinear and generalized linear models. In S. Ghosh & C. Rao (Eds.), *Handbook of statistics 13: Design and analysis of experiments* (pp. 437–475). Elsevier.
- Balliauw, M., Ongena, E., & Mulkens, S. (2020). Identifying factors affecting the value of advertisements on football clubs' and players' social media: A discrete choice analysis. *International Journal of Sports Marketing and Sponsorship*.
- Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 1281–1311.
- Basoalto, E., Hilton, R., & Knight, A. (2013). Factors affecting the efficacy of a vinegar trap for *Drosophila suzukii* (Diptera; Drosophilidae). *Journal of Applied Entomology*, 137(8), 561–570.
- Becerra, M., Balliauw, M., Goos, P., De Borger, B., Huyghe, B., & Truyts, T. (2023). Why do fans go to football games? A discrete choice analysis of

- ticket buyers' preferences. *International Journal of Sports Marketing & Sponsorship*. <https://doi.org/10.1108/IJSMS-05-2023-0093>
- Becerra, M., & Goos, P. (2021). Bayesian I-optimal designs for choice experiments with mixtures. *Chemometrics and Intelligent Laboratory Systems*, 217, 104395.
- Becerra, M., & Goos, P. (2023). Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables. *Food Quality and Preference*, 110, 104928.
- Bennett, J., & Blamey, R. (2001). *The choice modelling approach to environmental valuation*. Edward Elgar Publishing.
- Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7), 677–693.
- Bliemer, M. C., & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, 44(6), 720–734.
- Bliemer, M. C., & Rose, J. M. (2011). Experimental design influences on stated choice outputs: An empirical study in air travel choice. *Transportation Research Part A: Policy and Practice*, 45(1), 63–79.
- Bliemer, M. C., Rose, J. M., & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological*, 43(1), 19–35.
- Boonaert, E., Van Hoyweghen, K., Duguma Feyisa, A., Goos, P., & Maertens, M. (2023). Gendered fertility preferences and child schooling: Insights on the quantity-quality trade-off from Ethiopia. *Journal of Demographic Economics*. To appear.
- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324–345.
- Brent, R. P. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs: Prentice Hall.
- Burgess, L., & Street, D. J. (2005). Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference*, 134(1), 288–301.
- Cha, D. H., Landolt, P. J., & Adams, T. B. (2017). Effect of chemical ratios of a microbial-based feeding attractant on trap catch of *Drosophila suzukii* (Diptera: Drosophilidae). *Environmental Entomology*, 46(4), 907–915.
- Cornell, J. A. (1988). Analyzing data from mixture experiments containing process variables: A split-plot approach. *Journal of Quality Technology*, 20(1), 2–23.
- Cornell, J. A. (2002). *Experiments with mixtures*. Wiley.
- Cornell, J. A. (2011). *A primer on experiments with mixtures* (Vol. 854). John Wiley & Sons.
- Cornell, J. A., & Gorman, J. W. (1984). Fractional design plans for process variables in mixture experiments. *Journal of Quality Technology*, 16(1), 20–38.
- Courcoux, P., & Séménou, M. (1997a). Preference data analysis using a paired comparison model. *Food Quality and Preference*, 8(5-6), 353–358.

- Courcoux, P., & Séménou, M. (1997b). Une méthode de segmentation pour l'analyse de données issues de comparaisons par paires. *Revue de Statistique Appliquée*, 45(2), 59–69.
- Crabbe, M., Akinc, D., & Vandebroek, M. (2014). Fast algorithms to generate individualized designs for the mixed logit choice model. *Transportation Research Part B: Methodological*, 60, 1–15.
- Cruz-Esteban, S. (2021). It is not the color of the trap, but the color as a close-range stimulus inside the trap that increases capture of *Drosophila suzukii* and *Zaprionus indianus* (Diptera: Drosophilidae) in berry crops. *Crop Protection*, 141, 105449.
- Danthurebandara, V. M., Yu, J., & Vandebroek, M. (2011). Sequential choice designs to estimate the heterogeneity distribution of willingness-to-pay. *Quantitative Marketing and Economics*, 9, 429–448.
- David, H. A. (1963). *The method of paired comparisons* (Vol. 12). London.
- DeGroot, M. H. (2005). *Optimal statistical decisions* (Vol. 82). John Wiley & Sons.
- Eddelbuettel, D. (2013). *Seamless R and C++ integration with Rcpp* [ISBN 978-1-4614-6867-7]. Springer.
- Eddelbuettel, D., & Balamuta, J. J. (2018). Extending R with C++: A Brief Introduction to Rcpp. *The American Statistician*, 72(1), 28–36.
- Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8), 1–18. <https://www.jstatsoft.org/v40/i08/>
- Eddelbuettel, D., & Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71, 1054–1063.
- Feit, E. M., Feinberg, F. M., & Lenk, P. J. (2017). Bayesian analysis. *Advanced Methods for Modeling Markets*, 493–554.
- Fletcher Jr, R. J., Robertson, E. P., Wilcox, R. C., Reichert, B. E., Austin, J. D., & Kitchens, W. M. (2015). Affinity for natal environments by dispersers impacts reproduction and explains geographical structure of a highly mobile bird. *Proceedings of the Royal Society B: Biological Sciences*, 282(1814), 20151545.
- Fountain, M. T., Badiie, A., Hemer, S., Delgado, A., Mangan, M., Dowding, C., Davis, F., & Pearson, S. (2020). The use of light spectrum blocking films to reduce populations of *Drosophila suzukii* Matsumura in fruit crops. *Scientific Reports*, 10(1), 15358.
- Gabry, J., & Mahr, T. (2022). bayesplot: Plotting for Bayesian Models [R package version 1.10.0]. <https://mc-stan.org/bayesplot/>
- Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian Workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2), 389–402.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis, third edition*. Taylor & Francis.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 457–472.
- Goldfarb, H. B., Anderson-Cook, C. M., Borror, C. M., & Montgomery, D. C. (2004). Fraction of design space plots for assessing mixture and mixture-process designs. *Journal of Quality Technology*, 36(2), 169–179.

- Goos, P., Syafitri, U., Sartono, B., & Vazquez, A. (2020). A nonlinear multidimensional knapsack problem in the optimal design of mixture experiments. *European Journal of Operational Research*, 281(1), 201–221.
- Goos, P. (2022). The fish patty experiment: A strip-plot look. *Journal of Quality Technology*, 54(2), 236–248.
- Goos, P., Dens, N., De Pelsmacker, P., & Aleksandrovs, L. (2019). Using mixture-amount modeling to optimize the advertising media mix and quantify cross-media synergy for specific target groups. *Applied Stochastic Models in Business and Industry*, 35, 1228–1252.
- Goos, P., & Hamidouche, H. (2019). Choice models with mixtures: An application to a cocktail experiment. *Food Quality and Preference*, 77, 135–146.
- Goos, P., & Jones, B. (2011). *Optimal design of experiments: A case study approach*. John Wiley & Sons.
- Goos, P., Jones, B., & Syafitri, U. (2016). I-optimal design of mixture experiments. *Journal of the American Statistical Association*, 111(514), 899–911.
- Goos, P., & Syafitri, U. (2014). V-optimal mixture designs for the qth degree model. *Chemometrics and Intelligent Laboratory Systems*, 136, 173–178.
- Goos, P., Vermeulen, B., & Vandebroek, M. (2010). D-optimal conjoint choice designs with no-choice options for a nested logit model. *Journal of Statistical Planning and Inference*, 140(4), 851–861.
- Grasshoff, U., Großmann, H., Holling, H., & Schwabe, R. (2003). Optimal paired comparison designs for first-order interactions. *Statistics*, 37(5), 373–386.
- Haaijer, R., Kamakura, W. A., & Wedel, M. (2001). The 'no-choice' alternative in conjoint choice experiments. *International Journal of Market Research*, 43.
- Hamby, K. A., Kwok, R. S., Zalom, F. G., & Chiu, J. C. (2013). Integrating circadian activity and gene expression profiles to predict chronotoxicity of *Drosophila suzukii* response to insecticides. *PLOS ONE*, 8(7), e68472.
- Hamilton, N. E., & Ferry, M. (2018). ggtern: Ternary diagrams using ggplot2. *Journal of Statistical Software, Code Snippets*, 87(3), 1–17.
- Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools* [R package version 0.3.4]. <https://CRAN.R-project.org/package=purrr>
- Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 33(3), 307–317.
- Iglesias, L. E., Nyoike, T. W., & Liburd, O. E. (2014). Effect of trap design, bait type, and age on captures of *Drosophila suzukii* (Diptera: Drosophilidae) in berry crops. *Journal of Economic Entomology*, 107(4), 1508–1518.
- Jaffe, B. D., & Guédot, C. (2019). Vertical and temporal distribution of spotted-wing drosophila (*Drosophila suzukii*) and pollinators within cultivated raspberries. *Pest Management Science*, 75(8), 2188–2194.
- Johnson, R. M., & Orme, B. K. (1996). How many questions should you ask in choice-based conjoint studies. *Art Forum, Beaver Creek*, 1–23.
- Kessels, R., Goos, P., & Vandebroek, M. (2006). A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research*, 43(3), 409–419.
- Kessels, R., Jones, B., Goos, P., & Vandebroek, M. (2009). An efficient algorithm for constructing Bayesian optimal choice designs. *Journal of Business & Economic Statistics*, 27(2), 279–291.

- Kessels, R., Jones, B., Goos, P., & Vandebroek, M. (2011). The usefulness of Bayesian optimal designs for discrete choice experiments. *Applied Stochastic Models in Business and Industry*, 27(3), 173–188.
- Khademi, E., & Timmermans, H. (2012). Application of mixture-amount choice experiment for accumulated transport charges. *Procedia-Social and Behavioral Sciences*, 54, 483–492.
- Khademi, E., Timmermans, H., & Borgers, A. (2013). Traveler response to coexisting multiple pricing schemes: Results of elaborated mixture-amount experiment. *Transportation Research Record*, 2345(1), 63–73.
- Kowalski, S., Cornell, J. A., & Geoffrey Vining, G. (2000). A new model and class of designs for mixture experiments with process variables. *Communications in Statistics-Theory and Methods*, 29(9-10), 2255–2280.
- Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making: A user's guide. *Pharmacoconomics*, 26, 661–677.
- Lasa, R., Tadeo, E., Toledo-Hérnandez, R. A., Carmona, L., Lima, I., & Williams, T. (2017). Improved capture of Drosophila suzukii by a trap baited with two attractants in the same device. *PloS One*, 12(11), e0188350.
- Lee, J. C., Shearer, P. W., Barrantes, L. D., Beers, E. H., Burrack, H. J., Dalton, D. T., Dreves, A. J., Gut, L. J., Hamby, K. A., Haviland, D. R., et al. (2013). Trap designs for monitoring Drosophila suzukii (Diptera: Drosophilidae). *Environmental entomology*, 42(6), 1348–1355.
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, 100(9), 1989–2001.
- Little, C. M., Rizzato, A. R., Charbonneau, L., Chapman, T., & Hillier, N. K. (2019). Color preference of the spotted wing Drosophila, Drosophila suzukii. *Scientific Reports*, 9(1), 16051.
- Liu, H., Zhang, Z., & Grimm, K. J. (2016). Comparison of inverse Wishart and separation-strategy priors for Bayesian estimation of covariance parameter matrix in growth curve analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(3), 354–367.
- Lizin, S., Rousseau, S., Kessels, R., Meuldres, M., Pepermans, G., Speelman, S., Vandebroek, M., Van Den Broeck, G., Van Loo, E. J., & Verbeke, W. (2022). The state of the art of discrete choice experiments in food research. *Food Quality and Preference*, 102, 104678.
- Louviere, J. J., Pihlens, D., & Carson, R. (2011). Design of discrete choice experiments: A discussion of issues that matter in future applied research. *Journal of Choice Modelling*, 4(1), 1–8.
- Luyten, J., Kessels, R., Goos, P., & Beutels, P. (2015). Public preferences for prioritizing preventive and curative health care interventions: A discrete choice experiment. *Value in Health*, 18(2), 224–233.
- McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman; Hall/CRC.
- McFadden, D., & Train, K. E. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470.

- Melero, Y., Cornulier, T., Oliver, M. K., & Lambin, X. (2018). Ecological traps for large-scale invasive species control: Predicting settling rules by recolonising American mink post-culling. *Journal of Applied Ecology*, 55(4), 1769–1779.
- Meyer, R. K., & Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics*, 37(1), 60–69.
- Mitsui, H., Takahashi, K. H., & Kimura, M. T. (2006). Spatial distributions and clutch sizes of drosophila species ovipositing on cherry fruits of different stages. *Population Ecology*, 48(3), 233–237.
- Mortelmans, J., Casteels, H., Beliën, T., et al. (2012). Drosophila suzukii (Diptera: Drosophilidae): A pest species new to Belgium. *Belgian Journal of Zoology*, 142(2), 143–146.
- Otter, T. et al. (2022). Bayesian models. *Handbook of Market Research*, 719.
- Ozol-Godfrey, A., Anderson-Cook, C. M., & Montgomery, D. C. (2005). Fraction of design space plots for examining model robustness. *Journal of Quality Technology*, 37(3), 223–235.
- Piepel, G. F., Cooley, S. K., & Jones, B. (2005). Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm. *Quality Engineering*, 17(4), 579–594.
- Piepel, G. F., & Cornell, J. A. (1987). Designs for mixture-amount experiments. *Journal of Quality Technology*, 19(1), 11–28.
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Rehman, S. U., Paterson, A., & Piggott, J. R. (2007). Optimisation of flours for chapatti preparation using a mixture design. *Journal of the Science of Food and Agriculture*, 87(3), 425–430.
- Renkema, J. M., Buitenhuis, R., & Hallett, R. H. (2014). Optimizing trap design and trapping protocols for Drosophila suzukii (Diptera: Drosophilidae). *Journal of Economic Entomology*, 107(6), 2107–2118.
- Rice, K. B., Short, B. D., Jones, S. K., & Leskey, T. C. (2016). Behavioral Responses of Drosophila suzukii (Diptera: Drosophilidae) to Visual Stimuli Under Laboratory, Semifield, and Field Conditions. *Environmental Entomology*, 45(6), 1480–1488.
- Rodriguez, M., Jones, B., Borror, C. M., & Montgomery, D. C. (2010). Generating and assessing exact G-optimal designs. *Journal of Quality Technology*, 42(1), 3.
- Rossetti, T., Broaddus, A., Ruhl, M., & Daziano, R. (2023). Commuter preferences for a first-mile/last-mile microtransit service in the United States. *Transportation Research Part A: Policy and Practice*, 167, 103549.
- Rossi, P. E. (2019). *Bayesm: Bayesian inference for marketing/micro-econometrics* [R package version 3.1-4]. <https://CRAN.R-project.org/package=bayesm>
- Rossi, P. E., Allenby, G. M., & McCulloch, R. (2012). *Bayesian statistics and marketing*. John Wiley & Sons.
- Ruseckaite, A., Goos, P., & Fok, D. (2017). Bayesian D-optimal choice designs for mixtures. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(2), 363–386.

- Ryan, E. G., Drovandi, C. C., McGree, J. M., & Pettitt, A. N. (2016). A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1), 128–154.
- Scheffé, H. (1958). Experiments with mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2), 344–360.
- Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2), 235–251.
- Skaltsi, A., Marinopoulou, A., Poriazi, A., Petridis, D., & Papageorgiou, M. (2022). Development and optimization of gluten-free biscuits with carob flour and dry apple pomace. *Journal of Food Processing and Preservation*, 46(10), e15938.
- Smucker, B., Krzywinski, M., & Altman, N. (2018). Optimal experimental design. *Nature Methods*, 15(8), 559–560.
- Stan Development Team. (2022). *Stan user's guide*. Version 2.32. <https://mc-stan.org>
- Stan Development Team. (2023). RStan: The R interface to Stan [R package version 2.21.8]. <https://mc-stan.org/>
- Syafitri, U. D., Sartono, B., & Goos, P. (2015). I-Optimal Design of Mixture Experiments in the Presence of Ingredient Availability Constraints. *Journal of Quality Technology*, 47(3), 220–234. <https://doi.org/10.1080/00224065.2015.11918129>
- Tait, G., Mermer, S., Stockton, D., Lee, J., Avosani, S., Abrieux, A., Anfora, G., Beers, E., Biondi, A., Burrack, H., et al. (2021). *Drosophila suzukii* (Diptera: Drosophilidae): a decade of research towards a sustainable integrated pest management program. *Journal of Economic Entomology*, 114(5), 1950–1974.
- Thurstone, L. L. (1927). Psychophysical analysis. *The American Journal of Psychology*, 38(3), 368–389.
- Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., & Tuerlinckx, F. (2011). Visualizing distributions of covariance matrices. *Technical report, Department of Psychology, University of Leuven*.
- Torres, A. B., MacMillan, D. C., Skutsch, M., & Lovett, J. C. (2013). Payments for ecosystem services and rural development: Landowners' preferences and potential participation in western Mexico. *Ecosystem Services*, 6, 72–81.
- Train, K. E. (2001). A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit. *Department of Economics, University of California, Berkeley*.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.
- Vardakis, M., Goos, P., Adriaensen, F., & Matthysen, E. (2015). Discrete choice modelling of natal dispersal: 'Choosing' where to breed from a finite set of available areas. *Methods in Ecology and Evolution*, 6(9), 997–1006.
- Vermeulen, B., Goos, P., & Vandebroek, M. (2008). Models and optimal designs for conjoint choice experiments including a no-choice option. *International Journal of Research in Marketing*, 25(2), 94–103.
- Vojáček, O., Pecáková, I. et al. (2010). Comparison of discrete choice models for economic environmental research. *Prague Economic Papers*, 19(1), 35–53.

- Walsh, D. B., Bolda, M. P., Goodhue, R. E., Dreves, A. J., Lee, J., Bruck, D. J., Walton, V. M., O'Neal, S. D., & Zalom, F. G. (2011). *Drosophila suzukii* (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding its Geographic Range and Damage Potential. *Journal of Integrated Pest Management*, 2(1), G1–G7. <https://doi.org/10.1603/IPM10010>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wickham, H., Hester, J., & Chang, W. (2020). *devtools: Tools to make developing R packages easier* [R package version 2.3.2]. <https://CRAN.R-project.org/package=devtools>
- Yang, D., Timmermans, H., & Borgers, A. (2016). The prevalence of context-dependent adjustment of activity-travel patterns in energy conservation strategies: Results from a mixture-amount stated adaptation experiment. *Transportation*, 43(1), 79–100.
- Yu, J., Goos, P., & Vandebroek, M. (2010). Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B: Methodological*, 44(10), 1268–1289.
- Yu, J., Goos, P., & Vandebroek, M. (2011). Individually adapted sequential bayesian conjoint-choice designs in the presence of consumer heterogeneity. *International Journal of Research in Marketing*, 28(4), 378–388.
- Zahran, A., Anderson-Cook, C. M., & Myers, R. H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, 35(4), 377–386.
- Zijlstra, T., Goos, P., & Verhetsel, A. (2019). A mixture-amount stated preference study on the mobility budget. *Transportation Research Part A: Policy and Practice*, 126, 230–246.

Curriculum vitae

Mario Becerra holds a bachelor of science degree in Applied Mathematics and a master of science degree in Computer Science, both from Mexico's Autonomous Institute of Technology (ITAM). During his bachelor's program he chose electives related to statistics and computer science. His master's program aimed at teaching the fundamentals of computer science, together with a variety of practical electives that were from the Data Science master's program.

Before starting his PhD, he worked for five years in different companies and positions in Mexico performing data analysis. He built experience working with data, doing work that included data cleaning, data analysis, and data visualization. He also often worked in software building, programming, and version control.

In October 2019, he embarked on his doctoral studies in experimental design and choice modeling at the Faculty of Bioscience Engineering of KU Leuven.

Articles in international journals

- Mario Becerra and Peter Goos. “Bayesian I-optimal designs for choice experiments with mixtures.” *Chemometrics and Intelligent Laboratory Systems* (2021). DOI: 10.1016/j.chemolab.2021.104395
- Mario Becerra and Peter Goos. “Bayesian D-and I-optimal designs for choice experiments involving mixtures and process variables.” *Food Quality and Preference* (2023). DOI: 10.1016/j.foodqual.2023.104928
- Mario Becerra, Matteo Balliauw, Peter Goos, Bruno De Borger, Benjamin Huyghe, and Thomas Truyts. “Why do fans go to football games? A discrete choice analysis of ticket buyers’ preferences.” *International Journal of Sports Marketing and Sponsorship* (2023). DOI: 10.1108/IJSMS-05-2023-0093

Talks, seminars and posters

- Seminar titled *I-optimal versus D-optimal designs for choice experiments with mixtures* given online at the DOE-IT Seminars on Design of Experiments on June 19th, 2020.
- Talk titled *Bayesian I-optimal designs for choice experiments with mixtures* given at the ENBIS-21 Online Conference on September 14th, 2021.
- Poster titled *Bayesian optimal designs for choice experiments with mixtures* presented online at the Design and Analysis of Experiments 2021 Conference Series on October 14th, 2021.
- Talk titled *Bayesian D- and I-optimal designs for choice experiments with mixtures using a multinomial logit model*, given at the RSSB-21 Conference in Liège, Belgium on October 22nd, 2021.
- Seminar titled *Bayesian D- and I-optimal designs for choice experiments with mixtures and process variables using a multinomial logit model* given at the DOE-IT Seminars on Design of Experiments on May 10th, 2022 in Leuven, Belgium.
- Talk titled *Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables*, given at the International Choice Modelling Conference (ICMC) in Reykjavik, Iceland on May 25th, 2022.
- Poster titled *Bayesian optimal designs for choice experiments involving mixtures of ingredients and process variables* presented at the Bayesian Young Statisticians Meeting (BAYSM) in Montréal, Canada on June 22nd, 2022.
- Talk titled *Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables*, given at the 13th Model-Oriented Data Analysis and Optimum Design (mODa13) conference in Southampton, England on July 10th, 2023.
- Talk titled *Bayesian D- and I-optimal designs for choice experiments involving mixtures and process variables*, delivered at the Second International Workshop of the Scientific Research Network on Choice Modelling in Leuven, Belgium on September 14th, 2023.

FACULTY OF BIOSCIENCE ENGINEERING
DEPARTMENT OF BIOSYSTEMS
MECHATRONICS, BIOSTATISTICS AND SENSORS (MEBIOS)
Kasteelpark Arenberg 30
3001 Leuven

