Junior Management Science 7(3) (2022) 543-568

Junior Management Science

journal homepage: www.jums.academy

Pricing Sustainable Shipping of Coffee: Consumers' Preferences and Willingness to Pay for Emission Reductions and Offsets

Ulrich Bek

Technische Universität München

Abstract

The goal of this study is to explore whether decarbonization of maritime shipping and the full supply chain are valued in customer perception. Understanding consumers' preferences and willingness to pay (WTP) for sustainable maritime shipping of goods can provide opportunities to spread the costs of required sectorial changes. Decarbonization labels were compared to the Fairtrade and European organic label. A discrete choice experiment (DCE) with 299 participants was performed and supporting data was considered for an exhaustive description of preferences and WTP in the exemplary use case of filter coffee. The results indicate a significant WTP premium for all labels. On average, direct reductions of all supply chain emissions were valued at 2.82€ (all values per 500g of coffee) and are thus comparable in importance to the Fairtrade label estimated at 2.77€. Maritime shipping offsets, reductions and offsets for the full supply chain were valued at average premiums of 1.79€, 1.95€ and 1.89€. Organic labelling led to an average premium of 1.61€. A random parameter logit model with correlated parameters found significant preference heterogeneity across participants for all labels. Participants preferring whole bean to ground coffee did not significantly differ in their underlying preferences for the sustainability dimensions but in their price perception and effectively displayed higher WTP for all attributes. This study contributes to current research by providing a thorough measurement of preferences and WTP for emission reductions along the supply chain and is the first to assess offsets compared to direct reductions in a controlled setting for a common product use case.

Keywords: Sustainable maritime shipping; direct decarbonization; carbon offsets; discrete choice experiment; pricing.

1. Introduction

> "Everybody in the world benefits from shipping, yet few people realize it. We ship food, technology, medicines, and memories. As the world's population continues to grow, [...] efficient maritime transport has an essential role to play in growth and sustainable development. [...] Maritime transport is the backbone of global trade and the global economy."
>
> – Ban Ki-moon as United Nations Secretary-General (United Nations, 2016)

Maritime cargo shipping is a vital pillar of our globalized civilization. Relatively speaking, it is a comparably eco-friendly mode of transportation in terms of greenhouse gas (GHG) emissions per kilometer per ton of freight transported (Creutzig et al., 2014). In absolute terms however, the industry's mere size results in a substantial contribution to annual global emissions. In 2018 the sector's GHG emissions were estimated to add up to 1076 Mt carbon dioxide ($CO_{2)}$ equivalent, which corresponds to 2.89% of all annual global anthropogenic emissions (Faber & Kleijn, 2020). Mainly due to growing global trade, GHG emission increases of up to 50% are expected if no regulatory action is taken (Faber & Kleijn, 2020). But there is potential for a different scenario.

The industry is currently at a turning point. To reach the Paris Agreement (United Nations Treaty Collection, 2015) goals of limiting global warming to below 2 °C pressure on the relevant regulatory body, the International Maritime Organization (IMO), is increasing. In addition to institutional and regulatory pressures, market factors and social forces drive the IMO to act and lead the industry towards sustainable maritime transport (Serra & Fancello, 2020).

Psaraftis (2019) defines sustainable maritime transport as "striking the right balance between varied and potentially competing economic, social and environmental objectives".

DOI: https://doi.org/10.5282/jums/v7i3pp543-568

Maritime transport without causing any detrimental effects on the environment but lacking economic viability and consideration of social aspects does thus not qualify as fully sustainable.

This study primarily focuses on only one part of this definition. Only one sustainability criterion is in scope: GHG emissions. More specifically, it focuses on two categories of emissions, those caused by maritime shipping only, and those caused along the full product supply chain including maritime transport.

The GHG emission reduction efforts currently proposed by the IMO primarily revolve around three topics: technological measures (such as alternative fuels and renewable energy use), fleet-related operational and management measures (such as improvements in speed management and route planning), and market-based incentives that either "discourage the use of high-carbon fuels" or "encourage the adoption of low-carbon practices through incentives" (International Maritime Organization, 2020a, 2020b, 2020c; Serra & Fancello, 2020).

It will be challenging for the industry to comply with the expected regulatory changes estimated to cost the industry up to $60 billion USD per year (Pitt, 2017). Barriers to reach the IMO's emission goals include change aversion in the sector, investment-related risks, uncertainty about future regulations, information and time constraints, limits in the technological feasibility of measures proposed, as well as market-related issues and political obstacles (Serra & Fancello, 2020). In line with the beforementioned definition of sustainable shipping, the authors of a recommendable overview of these challenges conclude that, "the real challenge for the future is to succeed in effectively integrating environmental sustainability with economic sustainability and shipping needs" (Serra & Fancello, 2020).

An additional transitional approach to counteract the sector's environmental impact is to compensate for GHG emissions that cannot be prevented by funding measures to reduce emissions elsewhere (Meunier, Stoll, & Schoen, 2019). This approach called carbon offsetting is based on the idea that local emissions have global effects, and the sum of global emissions is what ultimately matters. It can thus be less effective to strive for zero emissions in one specific use case while neglecting other, more important emitters (Kollmuss, 2010). While the global market for voluntary and non-voluntary carbon offsets is growing rapidly, a lot of controversy is surrounding the approach for various reasons including ethical concerns and a lack of efficacy and efficiency of the offsetting measures taken (Day, 2021). The main critique on GHG offsets is summarized very well in a joint statement of multiple environmental organizations from 2006: "Purchasing offsets can be seen as an easy way out for governments, businesses and individuals to continue polluting without making changes to the way they do business or their behavior" (Canzi, Clough, & Kronick, 2006, p. 1). Direct reductions should thus usually be preferred to funding indirect reductions. Nonetheless, offsets can be a useful transition mechanism, but other policies can often result in faster and more efficient net reductions (Kollmuss, 2010).

This study contributes to aligning economic and sustainability objectives by addressing some topics that are currently (at least to some degree) neglected in literature. The three key issues in scope of this study are:

- understanding whether sustainable shipping is a relevant topic from a consumer perspective by estimating willingness to pay (WTP) for decarbonized maritime shipping and full decarbonization of supply chains,
- comparison of customer preferences for direct emission reductions with preferences for emission offsets, and
- comparison of the proposed (shipping) emission reduction labels and offset labels with established sustainability labels.

The general assumption that underlies this study is that in consumer perception goods shipped in a sustainable manner (causing less or no emissions) could be valued more than goods that were shipped conventionally. If this perceived value exists and influences decision behavior, this benefit can be quantified as a WTP. Estimating increases of WTP based on employing more environmentally friendly forms of shipping can enable innovations and business models built around this decarbonization and therefore act as a driver of change in the industry.

To test this assumption, customer preferences and WTP for compensation and reduction efforts are quantified in a discrete choice experiment (DCE) (Carson & Louviere, 2011) on a specific example use case: roasted ground and whole bean filter coffee. Coffee was chosen as the use case in scope for multiple reasons:

- Raw coffee is usually shipped by sea and then roasted near the target market. At least 80% of all German coffee demand is shipped by sea (Deutscher Kaffeeverband e.V. , 2020), as Europe lacks the climate to grow coffee[1].
- An estimated 61.8 % of the German population (14 years or older) consumes at least one serving of roasted filter coffee each day (Förster, 2020). Coffee is an omnipresent and frequently bought product. From a survey design perspective, a product like coffee that most participants use and should be acquainted with (in terms of common product features, prices, and packaging sizes) is preferable for accurate and valid results.
- Coffee suits the context of decarbonization well. As van Loo et al. (2015) point out, coffee can be considered a pioneering industry for sustainability certifica-

[1]The exception being one plantation in a suitable valley on the Canary Islands, which is geographically part of the African Plate but politically an autonomous region of the European Union.

tion schemes. It is a major industry for both environmental (e.g., organic) and social (e.g., Fairtrade) certification (Fairtrade International, 2021b). While carbon labels are still rather rare in the industry (van Loo et al., 2015) a niche of "Segel-Kaffee", sail-shipped coffee transported on sailing cargo ships does exist (Klein, 2021).

In addition to these theoretical advantages, a practical consideration has influenced the choice of coffee as a use case. This thesis was written in cooperation with the startup company CargoKite. With the goal of contributing to the decarbonization of the marine industry, CargoKite has developed a novel ship concept that makes it possible to use wind as the sole propulsion using a kite system. As this could allow them to achieve reliability and cruising speeds comparable to conventional cargo ships, their goal is to eventually compete with conventional container shipping at scale. This differentiates them from the few existing niche companies in the sector, for example EcoClipper (2021) that attempt to revive sailing cargo shipping on conventional, classical sailing ships in smaller scale niche markets.

An interesting beachhead market for initial CargoKite prototypes is (premium) coffee. With viable wind conditions on common coffee shipping routes and a stable global demand even in times of crisis (Deutscher Kaffeemarkt e.V, 2021; International Coffee Organization, 2021), (premium) coffee supply could be a viable first market to operate in, especially if consumers were willing to pay a moderate surcharge for products shipped sustainably to fund initial research and development cost. Thus, the practical relevance of this study lies in pricing the use case described to benefit CargoKite and other innovators driving decarbonization in comparable contexts.

An additional theoretical benefit of this study is that it provides first insights into whether emission offsets and direct reductions are valued differently by consumers. Furthermore, the study gives a quantified indication of the impact of creating a partially compared to a fully decarbonized supply chain, both for direct reductions and emission offsets. Influences of the Fairtrade (Fairtrade International, 2021b) and the European organic label (European Commission, 2021) on choice behavior are also measured in the study, mainly as a reference, to provide a realistic context for and indicate the plausibility of the above results. Of course, this research on the exemplary use case of coffee cannot provide a context-independent universally true value of WTP for sustainable (maritime) shipping, it does however provide a rigorously controlled exemplary estimate that can serve as a first indication for comparable settings of interest.

This study furthermore provides an applied example for performing a DCE with sequences of multinomial choices[2] (i.e., multiple discrete choices are made by each survey participant throughout different, statistically efficiently calculated choice sets and preferences are then derived from this data with a random parameter multinomial logit model), solely relying on[3] the free open-source statistics software R (R Core Team, 2021; RStudio Team, 2021). Both the generation of the choice design, as well as the analysis were performed using R and relevant packages (Croissant, 2020; Traets, Sanchez, & Vandebroek, 2020).

The study was performed online with a survey focused on the DCE. For context, individual specific variables (such as demographics, attitudes, comprehension measures and stated preferences) were also part of the survey.

In chapter two the presented work starts off with relevant theoretical background on the topics introduced above. A detailed record on the sample characteristics and experimental measures is provided in chapter three, followed by a thorough discussion of data analysis and the resulting findings in chapters four and five, respectively. After a brief discussion of limitations of this research in chapter six, the primary findings of this study are summarized in the conclusion (chapter seven).

2. Theoretical Background

The IMO has set ambitious goals to reduce their total annual GHG emissions by at least 50% compared to 2008 which requires reductions of approximately 85% per ship in operation (International Maritime Organization, 2020c). Reductions like these are only possible by broadly employing a mix of various measures. The IMO is currently in the process of defining and agreeing on short-, mid- and long-term measures, targeting 2023 to adopt and provide implementation schedules for the short-term measures that are deemed viable to reach their goals (Marine Environment Protection Committee [MEPC], 2018). This paper does not focus on the technical implications of implementing decarbonization in the industry. Nonetheless, to provide the necessary context to understand the issue at hand, the most important approaches to reduce emissions will be briefly summarized in the paragraph below. This summary is based on the work of Serra and Fancello (2020), a comprehensive overview of the measures as well as related challenges and opportunities.

As mentioned in the introduction, the main areas the industry focuses on to reduce shipping emissions are technological, operational (fleet-related), market-based and management measures. Most emissions are linked to fuel consumption. Today, the most common fuel is bunker oil (i.e., low quality diesel) causing relatively high emissions even in optimized modern marine engines. Thus, technological approaches revolve around alternative fuels with better emission performance, improvements in ship design efficiency, filtration of exhaust fumes and/or switching to alternative energy sources for propulsion like wind and solar energy. It

[2]This kind of study is sometimes referred to as a choice based conjoint study, as the term is ambiguous, this study will refer to methodology terms of Louviere's nomenclature instead (2011).

[3]Apart from the use of Qualtrics (2021), a service to host survey questions, all software used was open-source.

is currently uncertain which alternative will prove superior long term. Operational measures include better navigation and reductions in ship speed (also called slow steaming) and can strongly reduce fuel consumption and emissions. These measures can often be applied in short term but have the potential to conflict with other business goals e.g., by reducing overall transport capacity. Management and logistics-based approaches revolve around optimizing various specific processes related to maritime operations (such as berth allocation in ports) using simulations and mathematical modelling. Market-based measures are the third area of interest for decarbonization. They discourage the use of high emission technology and fuels (e.g., compulsory carbon taxes), or encourage the adoption of practices and technologies that result in lower emissions (e.g., subsidies for eco-friendly investments). Given the international complexity the shipping industry operates in, it is not considered likely to implement effective market-based mechanisms in the short but rather in the long term.

Serra and Fancello (2020) summarize multiple barriers towards implementing decarbonization measures at the required speed to reach the IMO's goals including:

- lacking technological maturity of some measures to function at scale;
- "chicken-and-egg" problems related to the changes required (e.g., ships using alternative fuels and the required infrastructure to provide these fuels lack operational viability without the other already in place);
- general risk aversion towards change and inertia of stakeholders;
- political obstacles (especially for market-based measures);
- economic barriers (such as required investment costs).

While many of these efficiency gains will also translate to better fuel economy, lower operational costs, and better economic performance (Raza, 2020), overall, compliance with the upcoming regulations the IMO proposes will be costly for the industry (Serra & Fancello, 2020). The question on how to share these investment costs across the ecosystems of those requiring shipping services remains unanswered (Egloff, Escudero, Sanders, Webster, & Zampelas, 2019).

The proposed study investigates an opportunity to share these costs by exploring whether consumers perceive value in decarbonized supply chains and are willing to pay a premium for sustainable maritime shipping. Comprehending sustainable shipping as customer value could be a solution or partial mitigator for the economic costs of change. This research focuses on quantifying the possible positive effects on WTP for products that were shipped in a sustainable manner in a DCE employing various forms of emission reduction and offset labels. Purely informational carbon footprint labels (e.g., showing the amount of carbon emissions caused without reducing or offsetting these emissions) are not within the scope of this study, while labels promoting offsets and reductions are. Thus e.g., the works of Akaichi, de Grauw, Darmon, and Revoredo-Giha (2016) and Steiner, Peschel, and Grebitus (2017) are considered out of scope due to carbon labelling with specific numbers (and their different use cases). Below, relevant customer preferences revealed in literature are explored resulting in the research question and hypotheses proposed.

2.1. Preferences and Willingness to Pay for (Shipping) Decarbonization

At the time of writing this thesis, to the best of the author's knowledge, no literature on WTP specifically for green maritime shipping practices on a product use case was available. Most research concerned with WTP for emission reductions and offsets is focused on other sectors such as residential energy (Streimikiene, Balezentis, Alisauskaite-Seskiene, Stankuniene, & Simanaviciene, 2019), aviation (Caputo, Nayga, & Scarpa, 2013; Schwirplies, Dütschke, Schleich, & Ziegler, 2019; Sonnenschein & Smedby, 2019), urban freight transport and road transportation (Lera-López, Faulin, & Sánchez, 2012; Polinori et al., 2018). Hence, selected, relevant results for WTP in different but possibly on some dimensions comparable settings are presented below. Delving into the various challenges of offset programs from a policy perspective exceeds the scope of this literature review, for an overview of offsetting programs, related risks, benefits, trends, qualification and certification issues please refer to Kollmuss (2010).

A study focusing on green transportation in general, not green maritime shipping was performed by Schniederjans and Starkey in 2014. Customer motivations, intentions and WTP to buy a t-shirt that was transported directly from "manufacturing [...] using a truck with energy efficient fuel" were explored applying the theory of planned behavior (Schniederjans & Starkey, 2014, p. 119). The study focuses on the motivational antecedents of WTP for green transportation. A frugal approach of directly asking for their participants' additional WTP for a "green transportation t-shirt" compared to a "general t-shirt" was employed, the results were analyzed descriptively based on multiple segmentation criteria and analyzed with structural equation analysis to model the impact of personal attitudes towards green transportation, perceived behavioral control and peer pressure on intention to purchase and WTP. Their results indicated significant effects of peer pressure and attitude on intention to purchase which in turn influenced WTP. There were significant gender differences, as males reported a lower additional mean WTP premium of \$4.06, 95% CI [\$3.58, \$4.55] than females, \$4.96, 95% CI [\$4.47, \$5.45][4]. Education, location, income, age and perceived behavioral control did not influence WTP significantly[5].

[4] Confidence intervals are reported in this overview if they were reported by the authors.

[5] Measured at $\alpha = 0,05$ significance level. A „significant" influence of age at $\alpha = 0,10$ is reported.

Polinori et al. (2018) have employed a similar approach on Italian university students in 2015, focusing on urban freight transportation and using a similar, vague "green transported t-shirt" label. The overall mean WTP premium for those who were willing to pay a premium was 4.86€ (SD = 3.3) per shirt. 155 out of 337 participants were not willing to pay a premium in an initial filter question. Females, as well as self-reported high-income as well as environmentally active participants and public transportation users were characterized by above average levels of WTP.

For more insights on the antecedents of individual WTP for (voluntary) carbon offsets, Tao, Duan, and Deng (2021) provide interesting insights based on modelling consumers' WTP for general voluntary carbon offsets using an extended theory of planned behavior. Specific knowledge about carbon offsetting and personal moral norms increased willingness to offset significantly by influencing attitude and perceived behavioral control. High-carbon consumers' willingness to offset was found to be strongly influenced by social status and social pressure concerns. Specific knowledge on carbon offsets and a detailed comprehension of the meaning of term carbon offsets cannot be assumed across the general population. Tao et al. (2021) reported low average knowledge of the concepts in their Chinese sample ($N = 905$). In the few previous studies published on the topic, average knowledge on carbon issues and carbon offsets in the US and Australia was also reported to be low (Polonsky, Garma, & Landreth Grau, 2011; Polonsky, Grau, & Garma, 2015).

Schwirplies, Dütschke, Schleich, and Ziegler (2017) examined the example of travel transportation, measuring the impact of individual factors and changes in framing on offset WTP. Participants with higher income, younger age and firmer environmental and politically social preferences were willing to pay more in the German sample ($N = 1005$). Individuals that believe in the efficacy of offsets for protecting the climate also show higher levels of WTP. Findings from Pleeging, van Exel, Burger, and Stavropoulos (2021) support this notion of the importance of efficacy beliefs, more specifically hope, as hopeful respondents from the Netherlands were more likely to pay more for emission reductions by sourcing green energy.

Besides individual factors, several external factors can improve the likelihood to voluntarily offset carbon emissions and/or increase the WTP for offsets. Blasch and Ohndorf (2015) differentiate between the WTP for offsets and the probability to offset in the first place. The first is predicted best by individual's internalized norms to avoid environmental degradation and partly on income. The latter is better explained by the expected social recognition for offsetting. Berger (2019) illustrate further framing effects on green products using signaling theory. Participants exhibited a higher WTP for green products when the product choice was public rather than private and the products signaled their sustainability clearly. Products that are costlier than their nongreen counterparts should thus be designed or labelled in a clearly recognizable way. Huber, Anderson, and Bernauer (2018, p. 235) reported that a combination of institutional signaling through publicly announced government policy (i.e., information about the Swiss government forcing industrial actors to offset) and group norm interventions e.g., "many of my friends are already voluntarily compensating their emissions" were most effective in persuading participants to express a willingness to offset and actual WTP. Similarly, matching of offsetting contributions by the travel provider, which could also be interpreted as a form of institutional signaling, also increased WTP in a different study Schwirplies et al. (2019).

Carbon label messaging that is framed as a gain-frame (e.g., "if you choose to offset your carbon emissions, you will be removing carbon from the atmosphere and helping to preserve our environment") results in significantly higher purchase intention of carbon offsetting products and significantly increased WTP, especially when combined with objective climate change information (e.g., "emission levels now exceed 400 parts per million, which has never occurred in the 800,000 years of recorded history") and objective carbon offsetting information (i.e., how the offsets are realized by whom) (Chi, Denton, & Gursoy, 2021, p. 5). Focus group discussions by Upham, Dendler, and Bleda (2011) resulted in the recommendation to use labels showcasing reduction efforts instead of specific emission values as they mean little to average consumers.

Regional projects and projects revolving around re-/afforestation were preferred to renewable energies and energy efficiency improvement projects by a German sample in 2019 (Schwirplies et al.). Ritchie, Kemperman, and Dolnicar (2021) reported a similar finding on the regionality of the projects for Australian air passengers but found that the freedom to choose a specific offsetting program does not increase WTP of air passengers. Baranzini, Borzykowski, and Carattini (2018) evaluated the acceptability of international in comparison to domestic reforestation offsets and found that emphasizing the cost-effectiveness of international reforestation programs can help reduce the preference for regional offsets. This preference might be contrary to rational economic incentives, as forestry projects are prone to risks that energy efficiency offsets do not inherit (e.g., reversal risks through natural disturbances such as insect outbreaks) (Galik & Jackson, 2009).

2.2. Preferences and Willingness to Pay for Sustainability Labels

Before diving deeper into the research question and experimental procedures undertaken, two important established labels also evaluated in this study will be briefly introduced: the Fairtrade label (Fairtrade International, 2021b) and the EG Bio Organic label (European Commission, 2021). Both labels are comparatively widespread for coffee and still grow in importance (Fairtrade International, 2021a; Willer, Trávníček, Meier, & Schlatter, 2021). An estimated 5% of all coffee sold in Germany in 2020 was certified as Fairtrade (Fairtrade International, 2021a) and approximately 4% of all coffee imported into the European Union in 2019 was certified as organic (Willer et al., 2021).

The main goal of the Fairtrade label is to "ensure fairer terms of trade between farmers and buyers, protect workers' rights, and provide the framework for producers to build thriving farms and organizations" (Fairtrade International, 2021b). Fairtrade is thus a predominantly economic and social label aimed at reducing international inequality and fighting poverty, its principles do however also incorporate certain environmental standards such as proper and safe management of chemicals, waste, soil and water resources. Fairtrade is widely recognized across industries and very present in the coffee industry. For example, 32,5% of all Fairtrade products sold were coffee products in 2019 (Fiedler, Frank, & Volland, 2020), thus it was chosen as an exemplary widely recognized label for social sustainability.

The second label that was part of the study is the European organic logo. It was introduced to give a coherent visual identity to European Union produced organic products. Several organic labels could have been chosen for this study, the EU organic logo was chosen because it is a mandatory label for all pre-packaged EU food products, produced and sold as organic within the EU (European Commission, 2021). It should thus be the most widely distributed and recognizable organic label for the sample. To give a detailed account of the detailed criteria for and criticism on both certifications exceeds the scope of this study. For reference please consult (Herrmann, 2015), for a recommendable overview over various German labels and their performance across multiple sustainability dimensions.

In contrast to the scarce literature on (maritime) transportation carbon reductions, multiple studies measuring the influence of established sustainability labels such as the Fairtrade and organic label (sometimes even of carbon neutral labels) on WTP in products use cases are available. The most relevant studies employing coffee as a use case and optimally estimating the impact of several labels in relation to each other will be briefly presented, to provide context for the results of this study.

Also, van Loo et al. (2015) have performed an experiment on coffee combining eye tracking with a discrete choice, evaluating the influence of visual attention on sustainability information. During their experiment WTP for coffee certified with USDA Organic, Fairtrade, Rainforest Alliance and a carbon footprint label was measured. The carbon footprint label, indicating "that the producer is reducing its carbon emissions", differs in content and is thus not objectively comparable to the carbon labels proposed for this study later. Based on a sample of 81 participants, the authors report the highest average WTP premium estimate of $1.16 per 12oz (∼340g) of coffee for the USDA Organic label, followed by the Rainforest Alliance and Fairtrade labels with $0.84 and $0.68, respectively. The carbon footprint label did not exert a significant influence on the choice behavior in their experiment.

A second study in a related domain was recently published by Birkenberg, Narjes, Weinmann, and Birner (2021) with results from a DCE performed in 2016 surveying a smaller sample of 80 German coffee bar visitors on their WTP for 250g packages of Arabica whole bean coffee. The packages varied on Fairtrade certification, a trust based direct trade without certification, as well as an offset based "carbon neutral" certification. The experiment resulted in a mean WTP premium of 1.77€ for carbon neutral offsetting, a mean WTP premium of 3.22€ for a non-certified direct trade claim and a mean WTP premium for Fairtrade certification of 4.30€ (all premia per 250g of coffee). The study furthermore contained insights from three focus group discussions on the topic. Interesting results included that while all groups mentioned fair trade relations as one of multiple important factors influencing their purchasing behavior, purely environmental aspects such as the carbon footprint had to be introduced by the moderator in every focus group and the participants showed little initial awareness of the environmental impact of coffee production. In an additional study focused on Fairtrade coffee Rotaris and Danielis (2011) reported an average WTP premium of 2.20€ for Fairtrade certification on a 250g package of coffee (beans or ground was not defined by the authors) based on a DCE of an Italian sample of 46 men and 89 women. Their own data and their review of previous studies in the domain indicated strong heterogeneity in premium prices across individuals. Factors explaining this heterogeneity included item type (e.g., ground coffee or brewed cups), age, gender, income, purchasing habits and individual views on ethical issues such as poverty and inequality.

Lastly, Lappeman, Orpwood, Russell, Zeller, and Jansson (2019) performed a DCE on Fairtrade coffee based on a South African sample ($N = 300$) incorporating personal values. They measured an overall mean WTP for Fairtrade of $1.22 per 250g of coffee (27% premium compared to the reference price). A cluster analysis based on the premium WTP was performed to analyze the influence of personal values. The only personal value showing a significant influence was humanitarianism i.e., the care for human welfare construct which was based on dimensions such as beliefs in equality and freedom. Besides this, few differences between the clusters were found, e.g., counter to the authors expectations, knowledge of Fairtrade did not differ between segments or influence WTP.

2.3. Research Question and Hypotheses

Considering this theoretical background, the research gaps this study addresses can be outlined. While there are multiple studies on WTP for established sustainability certification, labels promoting decarbonization are not yet as well understood. No study in the domain has been identified which evaluates consumers' perception of offsetting compared to direct reductions. Furthermore, this is the first study to evaluate the importance of sustainable maritime shipping to consumers. The main goal of this study is thus summarized in the following research question:

> **RQ:** What are consumers' preferences for price, emission offsets and reductions (for both maritime shipping only and the full supply chain)

as well as for Fairtrade and organic labelling in the context of 500g packages of filter coffee (ground or whole bean) and what is their resulting marginal WTP for these attributes?

This research question leads to the following testable hypotheses:

H1: On average, participants are willing to pay a premium for filter coffee that was shipped without causing net emissions, i.e., parameter estimates and the resulting marginal WTP premiums for direct emission reductions and offsets differ significantly from zero.

H2: On average, participants value decarbonization of the full supply chain more than decarbonization of parts of the supply chain (maritime shipping only), i.e., parameter estimates for the decarbonization of the full supply chain are significantly higher than the corresponding parameter estimates of maritime shipping only (for both, offsets and direct reductions).

H3: On average, participants attribute more value to direct emissions reductions compared to indirect offsets, i.e., parameter estimates for direct reductions are significantly higher than the corresponding parameter estimates for offsets (for both, maritime shipping only and decarbonization of the full supply chain).

H4: On average, participants that prefer ground or whole bean filter coffee have identical preferences for decarbonization, sustainability labels and price, i.e., relevant parameter estimates do not differ significantly between the participants prompted with whole bean and ground coffee.

2.4. Pricing Approach

An indirect choice based survey was chosen to test these hypothesis as it allows for a closer replication of realistic buying situations compared to direct estimation approaches (e.g., van Westendorp and contingent valuation) and is commonly viewed as the most suitable (but also one of the most complex and effortful approaches) to quantify customer preferences and WTP (Simon & Fassnacht, 2016). The process of identifying the most suitable category of pricing methods was guided by Simon and Fassnacht (2016). The DCE approach was chosen among different variants of choice-based approaches for multiple reasons. Firstly, it not only allows for relative preference statements between alternatives but also enables decisions to not purchase any goods, thus yielding more valid results (Simon & Fassnacht, 2016). Secondly, the method is recommended as it is characterized by reduced hypothetical and social desirability bias, especially for contexts including hypothetical and public goods relating to environmental topics (Drichoutis, Lusk, & Pappa, 2016; Norwood & Lusk, 2011).

Valuing the public good of mitigated CO_2 emissions as an attribute of a private product can further reduce hypothetical bias, as private products are less prone to this bias (Loomis, 2014). Contrary to popular belief and previous meta-analysis on the topic (Carson, Flores, Martin, & Wright, 1996; List & Gallet, 2001; Murphy, Allen, Stevens, & Weatherhead, 2005; Schmidt & Bijmolt, 2020) did not find evidence for the general, context-independent superiority of indirect valuation compared to direct approaches for private goods in terms measurement accuracy and hypothetical bias. The authors did however not differentiate between specific indirect and direct methods and modelling approaches, thus a properly executed and modelled DCE was still considered to be the most promising experimental estimation approach in the context. The recommendations by Hauber et al. (2016) and Hensher, Rose, and Greene (2015) informed all modelling decisions. The experimental approach was carefully chosen to potentially minimize hypothetical bias and is elaborated in more detail in the following chapters experimental procedures and data analysis.

3. Experimental Procedures

The following chapter will introduce the experimental measures taken throughout the experiment. The data collected for the main study consisted of the following measures presented in order of survey flow:

- preliminary in-store price comparison of ground and whole bean coffee
- stated buying habits of coffee products;
- socio-demographic characteristics;
- DCE;
- stated importance of relevant product aspects for coffee;
- an optional open question to address possible concerns about the CO_2 labels during the choice experiment;
- stated favorite coffee product with brand, product name, package size and average price (if such a product exists);
- a question battery to test the comprehension level of the difference between CO_2 compensation and direct reductions.

3.1. Preliminary In-Store Price Comparison

104 product offers from six stores in Munich, Germany, (two discounters, two supermarkets as well as one drug store and one organic store) were recorded from 24.02. to 31.03.2021. Product and brand names, grind (ground or whole bean), pricing, packaging size, Fairtrade and organic certifications were analyzed to inform the design of the choice experiment. The comparison showed that prices

varied widely between brands, roasts, and labels. The cheapest blends were sold at prices of 2.99€ to 3.49€ per 500g (e.g., "Ja! Kräftiger Röstkaffee") with the most expensive packages ranging up to approximately 15€ per 500g (e.g., "Martermühle Bio" product range). Whole bean coffee averaged at prices of 6.57€ per 500g while ground coffee averaged at 6.55€ per 500g, supporting the design choice to use the same price levels in both groups. The most common packaging size in the in-store comparison (500g) was chosen for the experiment.

3.2. Sample Characteristics and Recruitment

The German online survey was distributed from 12.04. to 29.04.2021. The resulting convenience sample was recruited through multiple channels including:

- the personal networks of CargoKite's founding team and the study's author,
- sharing on Facebook in various regional bulletin boards across Germany,
- untargeted sharing on LinkedIn and Xing.

Participating in the study was not incentivized. However, there is little evidence of positive or negative effects of incentives on response quality and sample composition in surveys (Singer & Ye, 2013). Of 462 total submissions, 436 contained data beyond the filter questions. A data check showed that some incomplete sets revealed signs of participants' fatigue (e.g., a single alternative was chosen throughout multiple choice sets). Submissions with complete choice data did not show any unfavorable patterns. As a precaution, submissions with incomplete choice data were excluded from the study. The resulting cleaned sample consisted of 299 submissions. This dropout rate of 31% might be considered higher than expectable for a survey of this length (Liu & Wronski, 2018) and will be discussed in the limitations section in more detail. The appropriateness of a sample's size depends on question format, choice task complexity, desired result accuracy, heterogeneity in the target population, respondent availability and whether subgroups separately analyzed (Bridges et al., 2011). Based on Johnson and Orme's (2003) sample size recommendations, the sample was appropriate in the context and given the DB-efficient design (discussed in detail the next section) a smaller sample size could have sufficed for valid results (Rose & Bliemer, 2013).

As shown in table 1, the sample was biased towards an above average education level and consisted of more females than males. Most of the sample stated to live in southern Germany. A filter excluded any participants that did not at least occasionally buy either ground or whole bean coffee. If both types were bought, participants were prompted with an additional question to confirm their most purchased type. Based on this data, they were assigned to one of two experimental groups, Ground Coffee or Whole Bean Coffee. Group demographics were mostly similar but relatively speaking group 1 (Ground Coffee) included more students and residents of Baden-Württemberg.

Both groups were presented with identical questions throughout the study, the exception being the labelling of the (otherwise identical) choice tasks either as ground or whole bean coffee to fit their preference. The inclusion of these two grind levels of coffee, two slightly different variants of one product, is in line with the study's goal to strike a balance between taking on a broad perspective on the research question, while also providing clearly specified, empirically grounded baseline information in the use case. Thus, the results can be useful as a reference for comparable contexts without compromising meaningfulness for the specific use case. Furthermore, by including both grinds, the recruitment of a bigger and more representative sample was possible, as ground roasted coffee has a 50% retail market share and whole beans accounts for an additional 37.3% of coffee sold in Germany in 2020 (Deutscher Kaffeemarkt e.V, 2021). Including two similar variants of the same product further gives a first indication of the robustness of the expected effects across slight changes in the context.

3.3. Design of the Discrete Choice Experiment

A DCE with multinomial choice questions and generic, non-labelled alternatives (except for the alternative specific no buy or status quo alternative) was performed.

The experiment was introduced using a direct translation of the cheap talk script proposed by Lusk (2003) to reduce potential bias in hypothetical valuation questions (cf. Appendix A). Clear descriptions of all attributes and levels are recommended to avoid comprehension issues and room for interpretation between subjects (Bridges et al., 2011). In the case of this study however, a balance had to be struck to give sufficient contextual information while avoiding unintended priming, social desirability and other confounding effects on participant behavior. The topic was introduced with concise information consisting of an explanation of the choice task and the cheap talk script to measure the decisions as neutral and as unbiased as possible. For example, emission levels were purposefully not explained to avoid emphasizing the differences between reductions and offsets, differences that participants might disregard in their natural choices. Inquiry into cognitive bias is not in scope of this study, for reference, Felser (2015) provides a straightforward overview of important psychological terms and biases mentioned in this study.

3.3.1. Presentation of the Experiment

The participants were presented with 14 predefined full profile choice sets in random order, each containing three alternatives including one no buy or status quo alternative. All profiles were introduced either as ground or whole bean filter coffee depending on group membership. The profiles were presented as simplistic 3D-rendered 500g coffee bags to mimic a realistic buying context (cf. Figure 1). To minimize unintended attention effects (cf. van Loo et al.'s (2015) study on the impact of visual attention on choice behavior),

Table 1: Socio-Demographic Sample Characteristics (in % of Segment, $N = 299$)

	Overall Sample	**Group 1: Ground Coffee**	**Group 2: Whole Bean Coffee**
Segment size	100.0	41.5	58.5
Gender			
Female	64.2	63.7	64.6
Male	35.1	36.3	34.3
Country of residence			
Germany	97.0	94.2	98.8
Austria	2.0	4.0	0.6
Other	1.0	1.6	0.6
County of residence			
Baden-Württemberg	54.8	42.7	63.4
Bavaria	29.4	37.9	23.4
Hesse	2.0	0.8	2.9
North Rhine-Westphalia	2.0	1.6	2.3
Other	8.0	12.9	4.5
Education level completed			
Secondary General School	2.7	1.6	3.4
Intermediate Secondary School	16.1	12.1	18.9
Grammar Schools (A-level)	25.8	24.2	26.9
Bachelor's degree	32.8	37.9	29.1
Master's degree	17.7	20.2	16.0
Other	3.3	1.6	4.6
Occupation			
Employees & civil servants	53.8	42.7	61.7
Student	32.1	42.7	24.6
Self-employed	4.3	2.4	5.7
Housewife or househusband	2.0	1.6	2.3
Retired	2.0	3.2	1.1
Other	3.0	3.2	2.9
Age (in years)			
Mean (*M*)	32.2	30.1	33.7
Standard Deviation (*SD*)	10.8	10.8	10.6

all labels were enlarged in same-sized boxes for better readability. A logo containing a globe with a transport route from America to Europe was added to emission attributes, to make sure all attributes were roughly equally visually appealing. Furthermore, all labels were colored in the same shade of green to ensure a similar contrast level. Figure 1 provides an original depiction of the German DCE as presented to participants, please refer to Appendix A for the original German survey questionnaire.

The order of the labels in the 3D-render was randomized at image creation. The order of the choice sets and the order of the profiles inside each choice set were randomized during the survey for each participant. The no buy (i.e., status quo) option remained on the right of the screen to reduce cognitive strain.

3.3.2. No Buy or Status Quo Alternative

A no buy alternative was added to create a more realistic setting (Rao, 2014), better mimic consumer choices and increase design efficiency (Brazell et al., 2006). The experiment was performed with a generic basic product excluding many value-adding elements (e.g., the package was not branded) and focuses on a new concept (carbon savings in shipping). Consequently, adding a no buy option was consid-

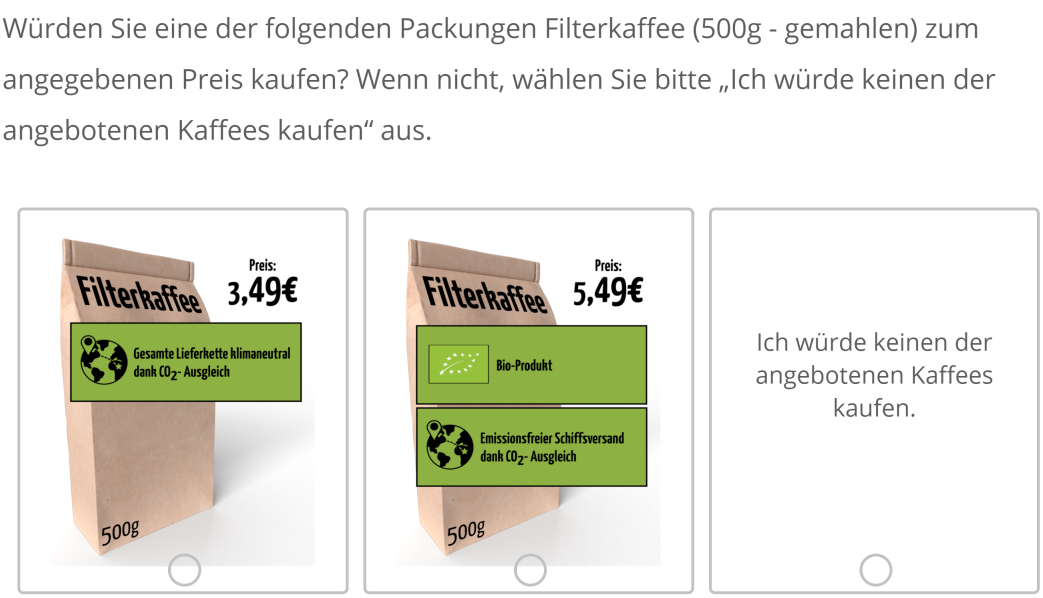

Figure 1: Depiction of a Survey Choice Set (Two Profiles & No Buy Option)

ered important to understand not only which alternatives and levels perform best in relative terms, but also whether they satisfy the participants minimum requirements for a (hypothetical) buying decision (Parker & Schrift, 2011). The disadvantages of adding a no buy alternative (e.g., lower information yield in sets where the no buy is chosen), were deemed less important than the above considerations, especially as most disadvantages can be mitigated by considering the no buy option correctly in the data analysis (Kamakura, Haaijer, & Wedel, 2001).

3.3.3. Attributes and Levels

The profiles differed on price (in € per 500g), emission labels, European organic certification, and Fairtrade labelling, as shown in table 2.

The continuous price levels were derived from the aforementioned in-store comparison and resemble common prices for ground and whole bean coffee. Both, the distance between price levels and the decimal places were kept constant between levels to avoid any confounding effects. Bliemer and Rose (2010) suggest that, while considering the risk of dominated alternatives, using a rather wide range is preferable to using a narrow range to avoid subjectively indistinguishable alternatives. As a result of these theoretical and practical considerations 4 steps of 2€ each from 3.49€ to 9.49€ were chosen as price levels. The grind was purposefully not added to the choice design as a level as it might not satisfy the basic DCE requirement of compensability (Gustafsson, Herrmann, & Huber, 2007). For example, a participant that does not own the equipment to grind coffee might never choose whole bean over bean, independent of the other attribute levels. The label level descriptions (organic, Fairtrade and carbon emissions) provide a simple, precise, and neutral account of the proposed value. Quantitative amounts of GHG emissions prevented/offset were not chosen as levels of the emission reduction attribute, assuming specific numbers mean little to average consumers (cf. Upham et al., 2011). Valuing the concepts of preventing and offsetting emissions is the goal of this study, while estimating WTP per ton of GHG emissions is not.

The categorical emissions attribute was defined in a balanced way, including four levels in total, two levels each for zero-emission maritime shipping and a climate neutral full supply chain, achieved by reductions and offsets, respectively. The number of levels for each attribute corresponds to the number of parameters that can be analyzed, therefore multiple levels were purposefully added as non-linear effects were expected for this attribute (Hess & Daly, 2014). The first carbon reduction level for zero-emission shipping was labelled with “Zero-emission maritime shipping thanks to wind energy” to describe the technology proposed by CargoKite as precise and neutral as possible. Although sail-shipped coffee is a somewhat more common term, it was not used to avoid evoking any emotional associations with sailing imagery. The second carbon reduction level was phrased “The full supply chain is climate-neutral thanks to emission avoidance with sustainable technologies”. Reducing emissions of the full supply chain to zero might not be technologically viable (yet), but the concept can already be described and the resulting WTP estimated. The corresponding offset levels were identically phrased, except for replacing the last part of the sentence with “thanks to CO_2-offsets”. The original German level descriptions can be found in Appendix B.

Only two specific labels (Fairtrade and organic) were added, to avoid inflating the design. Both labels are widely used on coffee and can be habitual buying criteria. Adding them provides a more realistic decision context. Furthermore, the labels’ results act as a reference to indicate the plausibility of the estimates in comparison with prior re-

Table 2: Choice Design with Profile Attributes and Levels

Level	Price	Organic	Fairtrade	Emission label
1	3.49€	No label	No label	No label
2	5.49€	Certified	Certified	Zero-emission maritime shipping thanks to wind energy.
3	7.49€	—	—	Zero-emission maritime shipping thanks to CO_2-offsets.
4	9.49€	—	—	Full supply chain is climate-neutral thanks to CO_2-offsets.
5	—	—	—	Full supply chain is climate-neutral thanks to emission avoidance with sustainable technologies.

Note: all prices in € per 500g of filter coffee

search. Adding the labels also allows for informative, comparative statements on the relevance and performance of the carbon labels relative to established labels for which market data and a richer body of existing research is available. The third reason for adding Fairtrade and organic labels was to provide an exemplary indication for the importance of social and production sustainability in contrast to GHG emission sustainability of maritime shipping.

3.3.4. Design Generation

The full list of possible profiles with these attributes and levels consists of 80 ($4 \times 2 \times 2 \times 5$) unique combinations. This full design was reduced to an efficient choice design using the modified Fedorov algorithm provided in the idefix package (Traets et al., 2020). The goal of this reduction was to achieve a set of tasks of feasible size that yields most information to estimate the parameters of interest. The algorithm optimizes the design for predefined parameter estimates while taking parameter uncertainty into account leading to a Bayesian d-efficient (so-called DB-efficient) design. Optimization is based on the DB-error i.e., the expected D-error minimized over the assumed prior distributions. This approach is in line with the recommendations of Traets et al. (2020) since designs optimizing for D-error are more sensitive to misspecifications of priors. Hensher et al. (2015) were referred to for theoretical guidance in defining the parameters and generating the design. The priors used for this optimization process were adopted from van Loo et al. (2015), who reported significant parameter coefficients for a price coefficient, organic and Fairtrade labels in a methodically and contextually comparable study. Using informative priors enables substantial improvements in the design's efficiency, and smaller ranges of the resulting parameter estimate confidence intervals in equally sized samples Carson and Louviere (2011). The parameter estimates for the emission labels were cautiously assumed to be zero. To maximize design robustness while taking participant fatigue into account, different set size specifications between nine and 16 choice sets per participant were estimated, following the common practice to include eight to 16 tasks per participant in a survey (Bridges et al., 2011). This iterative process resulted in a final design with a DB-Error of 2.56 consisting of 14 choice sets each containing two profiles and a no buy alternative.

3.4. Auxiliary Attitudinal and Knowledge Questions

To support the choice data, the stated importance of multiple possibly relevant attributes was evaluated using a 5-point Likert scale including a "no indication" answer option (cf. Appendix A). The DCE is focused on precise importance estimates of fewer product aspects, these ratings are intended to give a more exhaustive (although more frugal) overview over other influencing factors, their relative importance. The attributes and scales were inspired by Hasselbach and Roosen (2015) and modified by adding further buying criteria for coffee (e.g., brand, smell, origin, and taste) from a 2020 survey (Splendid Research) and attributes related to the research question like CO_2 impact. The order of the aspects was randomized between participants to avoid confounding positional effects.

To understand possible unintended influencing factors regarding the labelling or terminology used, an optional open question was added prompting participants to voice any concerns they might have towards emission-free or emission-compensation labels. This manipulation check monitors any negative attitudes towards emission reduction or compensation. For example, shipping emissions (although in absolute terms an important contributor to worldwide emissions Faber & Kleijn, 2020) might in relative terms be considered a small contributor in the lifecycle of coffee (Usva, Sinkko, Silvenius, Riipi, & Heusala, 2020). Psychological reactance could result from promoting it as a selling point. The measure was added after the choice experiment and the importance ratings to avoid unintended priming for concerns or consistency bias (Felser, 2015) in line with statements made.

An optional set of open questions was further added to evaluate stated product preference. Detailed information about participants' favorite (or habitually bought) coffee

product was requested. The product name and brand, followed by the packaging size and the average remembered price were only asked if they stated an existing preference. "No indication" options were included, packaging and price also contained an "unknown" option. These measures were intended to indicate habitual WTP and give a frugal indication of price interest and price knowledge.

As the last part of the survey, two quiz style batteries of variables were used to evaluate participants' comprehension of the difference between indirect emission compensations and direct emission reductions. Descriptions of situations were proposed, and participants had to choose whether these described CO_2 compensation or direct emission reductions, respectively. In total, six statements were presented for both tasks, two wrong statements and two statements that matched one of both definitions respectively. The order of the question batteries and the order of the items inside each battery were randomized for each participant to avoid confounding positional or learning effects. Prompting the participants to instead define the concepts and evaluate the answers manually was considered but refrained from to avoid straining participants and ensure objectivity of the results.

3.5. Data Analysis

There are multiple methods to conduct the statistical analysis of DCE data. This chapter will first introduce the software used to perform the analysis underlying this study. Secondly, it will give an overview over the choice model chosen and the assumptions that underly this model. Lastly, further complementary descriptive analysis undertaken in the context is introduced.

3.5.1. Software

The analysis was performed using R 4.10 "Camp Pontanezen" (R Core Team, 2021) in RStudio 1.4.1106 (RStudio Team, 2021). The packages tidyr (Wickham, 2021) and dplyr (Wickham, Romain, Henry, & Müller, 2021a) were used for general data transformation, epiDisplay (Chongsuvivatwong, 2018), ggplot2 (Wickham, Romain, Henry, & Müller, 2021b), texreg (Leifeld, 2013), Hmisc (Harrell, 2021) and skimr (Waring et al., 2021) were used for data visualization and summarization, idefix (Traets et al., 2020) and dfidx (Croissant, 2021) were used for generation of the choice design and formatting of the choice data and mlogit (Croissant, 2020) was used to perform the analysis and estimate all models. Finally, the rmarkdown (Allaire et al., 2021; Xie, Allaire, & Grolemund, 2018; Xie, Dervieux, & Riederer, 2020) package was used to create a reproducible workflows of all analysis performed. While advanced choice modelling is usually performed using paid specialized software, this work showcases that proper analysis adhering to scientific standards is feasible solely relying on free, easily accessible open-source software.

3.5.2. Analysis of the Choice Experiment with a Random Parameter Model

Multiple variations of models ranging from simpler multinomial fixed effects models to mixed logit models with various underlying assumptions were calculated for this study. The best and most suitable model to describe the data was a random parameter logit model (also called mixed logit model) taking the samples panel data structure into account and assuming all parameters to be normally distributed and correlated.

The random parameter logit model assumes that parameters vary from one individual to another and takes possible heterogeneity in the population into account (Croissant, 2020; Hauber et al., 2016; Hensher et al., 2015). As 14 repeated observations in the different choice sets were made, this longitudinal information was considered by assuming each individual can be described with a constant random parameter across all choice situations (Croissant, 2020; Hensher et al., 2015). The model considers the random parameters of individuals as random draws from a distribution whose parameters are estimated (Croissant, 2020; Hensher et al., 2015). It is necessary to define the properties of these random draws prior to performing the analysis. Pseudo-random draws based on Halton sequences were chosen, as these intelligent draws outperform truly random draws for simulation purposes in terms of stability and computational efficiency (Hensher et al., 2015). Multiple models with different ranges of draws from 100 to 2000 draws were estimated to confirm the stability of the final model (Hensher et al., 2015) and while not all models converged to completely stable solutions at first, neither the key results nor the parameter estimates changed drastically in later models. This lengthy estimation process is a commonly encountered problem for researchers employing the random parameter logit model (Hauber et al., 2016). The final model was created using 1000 (Halton) draws, as this (slightly) outperformed both models with more and less draws. Due to computational limitations only models with up to 2000 draws could be computed. In total, 20 models were computed ranging from 50 to 2000 draws and the best model based on the Log Likelihood value was chosen. The Akaike information criterion would have been an alternative measure of the goodness-of-fit of the model, correcting for complexity based on the number of parameters in the model (Field, 2013). As the same number of parameters were estimated in all models, the Log Likelihood value has similar informational value and was thus used to inform the model choice.

A Wald Chi-Squared Test from the mlogit package (Yves Croissant) was performed and concluded in a significant result, χ^2 (21, N = 299) = 653.11, p < .001. It tests the null hypothesis that the random effects are uncorrelated, the significant result thus indicates that the random parameters are correlated, supporting the choice of assuming correlations among the parameters. The resulting final model significantly outperformed all models assuming uncorrelated parameters in a Likelihood-ratio test comparing the correlated

model to the best uncorrelated model χ^2 (-21, N = 299) = 178.12, p < .001. The most suitable random parameter distribution for models with correlated parameters, a normal distribution, was chosen in modelling (Croissant, 2020).

To avoid confounding effects based on unobserved preference heterogeneity, all parameters were added to the model as random parameters (Hensher et al., 2015). In the resulting model each parameter is described with a mean[6], the estimated confidence intervals of this mean and a separate standard deviation of the distribution of the random parameters. This standard deviation, if significant, gives an indication of significant heterogeneity of the estimates in the sample. If heterogeneity is present in fixed effect models, this would lead to a reduced model fit but could in theory be handled through data segmentation (Hensher et al., 2015). For this study, an alternative approach could have been to create different fixed effect models for sample segments that exert similar choice behavior. In practice however, multiple reasons made the proposed model better suited in the context. Firstly, it would be unfeasible to measure and pick all the right segmentation criteria to explain this preference heterogeneity in a single study of this scale. Secondly, the given sample could become too small for meaningful and accurate results if divided into multiple segments. Thirdly, in the context of the study's scope a more realistic, general estimate is more meaningful than multiple segment-based estimates.

Another general benefit of mixed logit models is that they are not affected by the so-called independence of irrelevant alternatives property and thus provide a more complex but also more realistic model that does not assume that "the ratio of the choice probabilities is independent of the presence or absence of any other alternative in a choice set" (Hensher et al., 2015, p. 479). There was therefore no need to perform a Hausman-McFadden test for independence of irrelevant alternatives in this study (Hensher et al., 2015).

For a detailed mathematical derivation and description of the mathematical unconditional probabilities function that underlie the analysis performed, please refer to Croissant's (Croissant, 2020) for a concise overview, or to Train (2009) the original author of the theoretical content Croissant (2020) has applied in his work. The formulas used in the analysis are explained in much detail in the paper's chapter „5.1 Derivation of the model" more specifically in the section "Panel data", thus they were not added here to avoid redundancy.

The final model estimates the following parameters:

- The price attribute coded as a continuous variable.
- All other attributes, coded as dummy variables:
  - organic label;
  - Fairtrade label;
  - reduction of shipping emissions label;
  - reduction of full supply chain emissions label;
  - offset of shipping emissions label;
  - offset of full supply chain emission label.
- The no buy or status quo option was added to the analysis as an additional dummy variable to avoid a possibly lower model and predictive fit for the other (linear) attributes (cf. Kamakura et al. (2001) for more information).
- Lastly, interaction effects between group membership of group 1 (whole bean coffee) and all other parameters were added. These estimates reveal differences between this subsegment of the sample and the rest of the sample (if applicable).
- The intercept was omitted as the alternatives were not labelled (except for the No buy alternative that had a separate parameter) and an estimate of alternative-specific effects was thus not required.

The resulting WTP and WTP confidence intervals for the various labels in the WTP space was estimated based on the random parameter's marginal utility for each participant, dividing it by the corresponding individuals price coefficient estimate. Multiple WTP mean estimation techniques, e.g., the delta or Krinsky Robb methods as proposed by Hole (2007), were also considered, but given the availability of all required individual specific estimates the resulting WTP distribution could be directly calculated and the confidence intervals and properties of the resulting distribution were reported.

Several of the initial hypothesis can be answered by testing for equality of the parameter coefficients resulting from the model discussed above. The appropriate Z-tests were used to test for significant differences between these model coefficients (Paternoster, Brame, Mazerolle, & Piquero, 1998).

3.5.3. Further Analysis to Support the Discrete Choice Experiment

Further variables were analyzed to provide context for the results of the study, the following demographic data was analyzed descriptively to characterize the sample overall and divided by the two groups whole bean and ground coffee:

- age;
- gender;
- country of residence;
- county of residence in Germany (if applicable);
- education level;
- occupation;
- price and preferred package size of their favorite coffee.

[6] This mean is identical with the median, as the random parameter distribution is symmetric.

Furthermore, a manipulation check to monitor possible concerns that could cause psychological reactance towards the labels (Felser, 2015) was introduced. The qualitative data resulting from this optional open question was analyzed in a structured manner, guided by the Grounded Theory approach (Corbin & Strauss, 2015), to derive bottom-up descriptive categorical codes based on the data.

Comprehension of differences between offsets and reductions was analyzed based on the comprehension question battery data, descriptively, overall and by group, for offsets and reductions respectively. A score was calculated by adding a point for each statement matched correctly with the definition in each task. If a statement was matched falsely in one of the tasks, all points were lost for this task. This results in a scale from 0 (either no statement was correctly matched, or a mistake was made) to 2 (all statements were correctly matched without a mismatch). This mode of scoring was chosen as it allows for meaningful interpretations of the mean, which equals to the average number of tasks completed correctly. Differences between scores of groups of individuals were analyzed using non-parametric tests.

The stated Likert scale importance ratings for coffee attributes were treated as ordinal variables and analyzed descriptively using mean ranks, rank standard deviations and histograms. Correlations between the importance ratings, age, sex, and group membership were calculated using a Spearman's Rho correlation matrix to identify associated factors and reveal possible group differences in more detail.

4. Results

In the following chapter the study's results are presented starting with the analysis of stated importance ratings, followed by brief analysis of the in-store price comparison and learnings from habitually bought coffee in the sample. Conclusions from the manipulation check for greenwashing and the results of the comprehension check for offsets and reductions are presented. In this supporting context, the results of the DCE are introduced. Finally, in the following chapters five and six, these results are reviewed and discussed in light of existing literature and challenged in consideration of the study's limitations.

4.1. Stated Importance of Attributes and Consumption Preferences

The self-stated importance ratings of coffee attributes are summarized in Table 3, frequency bar charts for these attributes used for visual inspection can be found in Appendix C.

The main takeaways from this data were that taste seems to be the most important criterion for coffee overall, it was on average ranked much higher than the other criteria as "important" to "very important". Brands were rated as least important by the sample. The product attributes in scope of this study were ranked at medium importance. In general, the validity of this rank order should not be overestimated as the mean rank standard deviations are rather high indicating heterogeneity of preference in the sample. Plotting the sustainability ratings, all label histograms have a positive skew (most individuals rated them as important), with individual rating ranging from irrelevant to very important (cf. appendices C5, C7 and C8).

4.1.1. Exploratory Correlational Analysis of Stated Importance Ratings

The ranking data was analyzed exploratively in a correlational analysis employing a Spearman's Rho correlation matrix from the package Hmisc (Harrell, 2021). Insights from this analysis were very coherent with expectations and revealed multiple small differences in preference between the groups for whole bean and ground coffee. Dimensions relating directly to sustainability of coffee were correlated with medium effect sizes. For example, stated importance of the CO_2 footprint correlated with the stated importance of the organic label, $r_s = .41$, $p<.001$, $n = 293$. It also correlated with the rating of Fairtrade certification, $r_s = .43$, $p<.001$, $n = 292$, and the importance of product origin, $r_s = .33$, $p<.001$, $n = 288$. Interest in any sustainability criterion thus seems to be associated with interest in other sustainability criteria. When analyzing group membership effects, small differences are revealed. Group membership did however not correlate significantly with the importance of any sustainability dimensions[7] or with the reported gender.

Group 2 (whole bean coffee) did however slightly differ from group 1 (ground coffee) in age, importance of price and dimensions directly related to coffee quality:

- Group 2 (whole bean) participants were (significantly) older (weak effect) than participants in group 1: $r_s = .21$, $p < .01$, $n = 286$. As a reference, the mean age in group 1 was $M_1 = 30.1$ compared to $M_2 = 33.7$ in group 2 with standard deviations of $SD_1 = 10.8$ and $SD_2 = 10.6$ respectively.
- The second group's stated importance of price was (significantly) lower (weak effect) compared to group 1: $r_s = -.14$, $p = .013$, $n = 294$)
- Smell was rated as (significantly) more important (weak effect) in group 2 compared to group 1: $r_s = .23$, $p < .001$, $n = 291$.
- Taste was rated as (significantly) more important (weak effect) in group 2 compared to group 1: $r_s = .17$, $p<.001$, $n = 289$.

[7] Importance of Fairtrade label not correlated with group 2 membership: $r_s = -.02$, $p = .73$, $n = 293$ Importance of CO_2 footprint not correlated with group 2 membership: $r_s = -.02$, $p = .71$, $n = 291$ Importance of organic label not correlated with group 2 membership: $r_s = -.02$, $p = .79$, $n = 292$ Importance of product origin not correlated with group 2 membership: $r_s =$, 11, $p = .07$, $n = 288$ Furthermore, no correlation of gender with group 2 membership: $r_s = .02$, $p = .77$, $n = 297$

Table 3: Importance Ratings for Coffee Attributes in the Sample

	Rank order	Mean rank (SD)	*n*
Importance of:			
Taste	1	4.65 (0.74)	289
Digestibility	2	4.01 (1.16)	292
Previous product experiences	3	3.97 (1.07)	292
Smell	4	3.96 (1.07)	291
Fairtrade certification	5	3.62 (1.10)	293
Price	6	3.45 (0.98)	294
Carbon footprint	7	3.41 (1.09)	291
Organic certification	8	3.35 (1.10)	292
Type of roast	9	3.34 (1.24)	289
Origin	10	3.15 (1.15)	288
Convenience	11	2.83 (1.27)	282
Packaging	12	2.71 (1.12)	294
Brand	13	2.44 (1.17)	294

- The type of roast was also rated as (significantly) more important (weak effect) in group 2 compared to group 1: $r_s = .22$, $p<.0001$, $n = 289$.
- Coherent with expectations, as an additional step (grind) is needed to brew whole bean coffee, group 2 accredited (significantly) less importance to convenience (small effect) than group 1: $r_s = -.16$, $p<.001$, $n = 282$.
- Lastly, digestibility was also rated as slightly (but significantly) more important in group 2 than in group 1: $r_s = .15$, $p = .013$, $n = 292$.

4.1.2. Stated Habitual Product Preferences

45% of the 299 participants stated a favorite or habitually bought coffee product, with 136 participants indicating that they have no favorite coffee and 33 participants skipping the question. 69% of these 136 participants gave both a packaging size and a price estimate giving frugal indications of price interest and knowledge. Thus, the following data (see Table 4) describes only a subset of the sample but could give an indication of possible differences between the groups, nonetheless.

Welch Two Sample t-tests were performed to test if the two groups differ significantly on their mean price per kg and mean package size in g. While the reported price per kg did not differ significantly between the groups $t(81.57) = 1.60$, $p = .11$, the average package size did, $t(108.35) = 6.80$, $p<.001$, indicating that the whole bean coffee buyers were used to significantly larger packaging sizes than the ground bean buyers. As the chosen packaging size of 500g for the choice experiment represents a compromise between size expectations for both groups (and lies within a standard deviation for both), it proved to be a reasonable and realistic choice for the experiment. Qualitatively, few brands were cited multiple times (e.g., Lavazza, but also Darboven, Hochland, Mövenpick and Tchibo) and a varied mixture of local roasters, discounter home brands, online shops, and organic specialty coffees was mentioned hinting at diverse preferences. As a conclusion, stated preferences about favorite coffee products gave first indications of existing diversity of brand preference, price knowledge and interest in the sample.

4.2. Results from the In-Store Price Comparison

The sample of coffee packages was too small for meaningful differentiation of the effects of organic and Fairtrade certification in consideration of interaction effects as the coffees were often accredited with multiple labels. Controlling for confounding variables such as brand and roast would not have been possible given the sample's limited size. As a first indication of the magnitude of expected effects, on average the 35 certified products with at least one label were sold at higher prices ($M = 8.32$€ , $SD = 3.32$€) than the 70 products without organic or Fairtrade certification ($M = 5.69$€ , $SD = 2.75$€). A Two Sample Welch-test was computed to test whether this difference is statistically significant while accounting for the unequal sample sizes of the two groups and found a significant difference between the prices of certified and uncertified products, $t(56) = 4.02$, $p<.001$.

4.3. Qualitative Analysis of Reported Concerns

As previously discussed, a manipulation check for unintended reactance towards the emission-free shipping labels was added. 60 individuals voiced concerns, 30 were members of groups 1 and 2 respectively. Some submissions contained information relating to multiple analysis categories. Two main categories of concerns were voiced: insufficient description and the credibility of the statements. Some concerns were mentioned that did not relate directly to the issues at hand (e.g., "wind power plants also impact our ecosystem"[8]) or were hard to allocate due to insufficient informa-

[8]Exemplary original comments were translated trying to minimize changes in style or meaning.

Table 4: Description of Favorite* Coffee Products by Segment (n = 107)

	Overall Sample	**Group 1: Ground Bean**	**Group 2: Whole Bean**
Segment size in %	100	36	64
Package size in g:			
Mean	688	442	830
Standard Deviation	389	202	401
Price per kg in € :			
Mean	18.1	16.2	19.1
Standard Deviation	9.3	8.73	9.44

* The sample indicated their favourite or habitually bought coffee.

tion provided by the participants. The data was structured inductively, and not fitted onto an existing model resulting in the partially ordered overview of concerns provided in Table 5.

The first category "insufficient descriptions" can be further specified with three issues that emerged multiple times:

- lacking general comprehension of the meaning of the labels (e.g., "honestly, [I] cannot imagine what is meant and thus not assess the importance");
- lacking transparency on specific aspects of the labels and wishes for clarification of distinct aspects (e.g.,"[the label] suggests transport on a sailing ship (if that is the case: concrete description + certification! This way higher prices would be by all means acceptable");
- missing clarity of the logical link between powered by wind energy and shipping was also expressed by some participants (e.g., "[I] don't understand how something can be shipped by wind power [..]"), although this is also a specific transparency issue, as it was voiced by various participants it is presented as an additional subcategory. The concept of sailing might not be known to all participants, counter to the researcher's assumption.

The credibility concerns are specified further through multiple subcategories:

- concerns with the general credibility of certification and distrust towards labelling schemes (e.g., "[I] have little faith left in certifications and labels");
- greenwashing concerns (e.g., "If it is called emission-free but then restricted, this smells like cheating/ greenwashing"), concerns like these are known to cause consumer behavior that is opposed to the goals of affected companies or organizations (de Jong, Harkink, & Barth, 2018);
- uncertainty about the relevance of shipping compared to the total carbon footprint of coffee in the context of the product lifecycle (e.g., "I don't know how big the total effect is: E.g., shipped emission-free thanks to wind power might equal a 50% reduction of carbon in the whole value chain?");
- general disapproval of carbon offsetting practices (e.g., "I feel that offsetting emissions with other sustainability projects is sometimes an excuse not to deal with them in the product lifecycle";
- doubts on the technical feasibility of shipping without emissions at our current level of technology or given the current state of the industry (e.g., "Global transport of goods is not yet possible (completely) emission-free").

The main goal of this manipulation check was to assess whether a majority of participants showed reactance to the labels presented because of greenwashing concerns or failed to understand the tasks. However only 2% of the sample did report greenwashing concerns and only 2% of the sample reported confusion about the labels. Participants voicing these more serious concerns were not excluded to reflect the sample's preference diversity. Further questions that arose because of the purposefully frugal description of the context were expected. With 7% of the sample voicing slight needs for more clarification, this is certainly a valid point to consider in further research but given the general focus of this research the labels seem to have been described comprehensively.

4.4. Comprehension of Reduction and Offsets

Comprehension of the differences between reductions and offsets was tested with two quiz-style batteries. 94% of the sample took part in this quiz located as the last part of the survey. Of these 281 participants, only 16% were able to identify all statements and assign them to direct reductions or carbon offsets correctly.

Two variables were calculated to describe the comprehension of offsets and the comprehension of reductions, respectively. The sample scored higher when assigning examples to the direct reductions ($M = 1.14$, $SD = 0.83$)

Table 5: Frequency of Concerns Voiced in the Sample ($N = 299$)

Description of concerns on emission-free shipping	Frequency count	Percentage of sample
Total count of concerns voiced	60	20%
Insufficient description of the context:	22	7%
Lacking transparency on specific label aspects	8	3%
Lacking general comprehension of label meaning	7	2%
Unclear link between shipping and wind power	7	2%
Credibility concerns:	32	11%
General credibility of labels	11	4%
Greenwashing* i.e., marketing vs. altruistic motives	6	2%
Relevance of shipping vs. total carbon footprint	6	2%
General disapproval of carbon offsets	4	1%
Technical feasibility of shipping without emissions	4	1%
Unclear and unrelated concerns:	10	3%
Unclear information provided by participants	5	2%
General feedback without mentioning concerns	3	1%
Attitudes on non-related issues	2	1%

*Greenwashing describes making „people think that [an organization] is concerned about the environment, even if its real business [...] harms the environment" (Oxford University Press, 2021))

than to carbon offsets ($M = 0.65$, $SD = 0.8$). As the two variables did not visually resemble a normal distribution, a paired Wilcoxon signed-ranks test was performed to test if scores of both tasks differed significantly in their central tendency. Indeed, the participants' comprehension score for direct reductions was significantly higher compared to the comprehension score of carbon offsets $T = 9386$, $p<.001$. Two independent Wilcoxon rank-sum tests showed no significant differences in performance between groups one and two in terms of comprehensions of reductions ($W = 9309$, $p = 0.58$) and offset comprehension ($W = 10042$, $p = 0.53$). These results indicate that most participants did not have a perfect understanding of the definitions of the terms and that offsets were harder to understand than direct reductions.

4.5. Modelling Consumer Preferences using Random Parameter Logit

Table 6 shows an overview of the results of the final model summarized on the next page. In total, eight parameters and eight corresponding interaction effects were estimated. The random parameter logit model estimates multiple metrics for each parameter. The mean values of the parameters (β_p) describe the average part worth estimate across the population of respondents (Chapman & Feit, 2019). The standard error of the mean indicates the average deviation of this parameter estimate from the true population average value and thus measures the precision of this estimate. The next column shows the standard deviation of this mean (SD_p), signaling how the parameters varied across the sample's population between individuals (Chapman & Feit, 2019). The corresponding standard error of this estimate is provided. The standard deviation is a measure for the existence of homogeneity or heterogeneity of preferences between individuals belonging to the sample's population.

The significance levels of all estimates are shown by the asterisks (cf. legend below the table). The Significance of the mean estimates (β_P) indicates whether they deviate statistically significantly from zero. Significant estimates thus exerted a significant impact on the choices made. Significance of the standard deviation (SD_P) indicates that the coefficients vary significantly in the population and that a mixed logit model provides a significantly better representation of the given choice context than a comparable multinominal fixed effects model assuming no preference heterogeneity could. The size of all estimates needs to be interpreted in relation to the other parameter estimates in the model. To enable a more applicable interpretation in absolute terms (and comparability to other contexts), WTP values based on the parameter in scope and the price coefficient estimate were calculated. The resulting WTP values are presented, visualized and discussed in chapter 4.5.1.

The model shows that all parameter estimates have a statistically significant influence on the choices made by the participants. All labels had a positive impact on choice. If they were present, all other things equal, participants were more likely to choose the given alternative.

In relative terms, Fairtrade certification ($M = 3.69$, $SE = 0.09$) and the reduction of all supply chain emissions ($M =$

Table 6: Random Parameter Logit Model with Coefficient Estimates ($N = 299$)

Parameter	**Mean (SE)**		**Standard Deviation (SE)**	
Fairtrade label	3.69***	(0.09)	2.75***	(0.18)
Organic label	2.20***	(0.17)	1.86***	(0.16)
Offset maritime emissions	2.50***	(0.21)	1.84***	(0.21)
Offset full supply chain	2.73***	(0.30)	2.14***	(0.26)
Reduction maritime emissions	2.61***	(0.31)	1.97***	(0.25)
Reduction full supply chain	3.67***	(0.35)	2.06***	(0.28)
Price coefficient	-1.33***	(0.09)	1.02***	(0.07)
No buy coefficient	-3.08***	(0.46)	6.85***	(0.42)
Interactions terms with Group 2: Whole Bean				
Fairtrade label x G2	0.13	(0.21)		
Organic label x G2	-0.12	(0.09)		
Offset maritime emissions x G2	-0.44	(0.29)		
Offset of full supply chain x G2	-0.23	(0.33)		
Reduction maritime emissions x G2	-0.39	(0.34)		
Reduction full supply chain x G2	-0.55	(0.36)		
Price coefficient x G2	0.62***	(0.09)		
No buy coefficient x G2	2.00***	(0.49)		
Akaike's Information Criterion				4882.04
Log Likelihood				-2389
Num. obs.				4186
K				3

Note: $***p < .001$; $**p < .01$; $*p < .05$

3.67, $SE = 0.35$) were similar in strength and were the most important predictors of choice behavior. The coefficients for decarbonization of maritime shipping with reductions ($M = 2.61$, $SE = 0.31$) or offsets ($M = 2.50$, $SE = 0.21$) as well as the coefficient for offsetting the full supply chain ($M = 2.73$, $SE = 0.30$) were similar in strength, but influenced choice behavior less than the coefficients for Fairtrade and full reductions. The organic label ($M = 2.20$, $SE = 0.17$) had the relatively speaking smallest significant effect on the choices made. In line with expectations, the price coefficient was negative ($M = -1.33$, $SE = 0.09$), indicating that participants preferred smaller prices (all other things being equal) to higher prices. On average, the no buy option was also negative in sign ($M = -3.08$, $SE = 0.46$), indicating participants preferred choosing a coffee to not choosing a coffee across all sets. The definition of the priors (cf. chapter 3.3.4) has resulted in a noticeable impact on model precision, as the standard errors for parameters with predefined priors were comparatively smaller.

Preference heterogeneity was present for all labels. Their estimates had significant standard deviations ranging from 1.84 to 2.14 SD, except for Fairtrade with even larger deviation of $SD = 2.75$, $SE = 0.18$. The labels' influences were heterogenous across individuals, emphasizing that point-based estimates (e.g., using a multinomial model) would be insufficient to model these diverse preferences.

The second part of the table shows the added interaction terms of group membership with all eight parameter estimates to reveal possible effects correlated with the consumption of whole bean in comparison to ground coffee. None of the label estimates became significant, but both the price ($M = 0.62$, $SE = 0.09$) and no buy ($M = 2.00$ $SE = 0.49$) effects were significant and positive. The non-significant label interaction effects show that whole bean consumers did not differ significantly from ground bean consumers in their preferences for the sustainability labels. The positive price interaction term results in an overall price coefficient closer to zero for the segment, indicating that whole bean consumers were less reluctant to choose higher-priced coffee in the experiment. The positive, significant no buy estimate shows that the segment was on average more likely to choose the no buy alternative across all choice sets. This result signals that group 2 was less likely to compromise on their coffee choices if no coffee suited their preferences in a choice set.

The chosen random parameter logit model with correlated parameters also models the correlations between attribute levels. The focus of this research was not concerned with these correlations, they thus not discussed in detail in

this work. As a short summary of the findings, sustainability and decarbonization labels were significantly correlated with each other, indicating that participants whose choices were influenced by one label were also influenced by other labels. The price coefficient was not significantly correlated to any of the label coefficients.

4.5.1. Willingness to Pay Resulting from the Random Parameter Logit Model

The resulting WTP premiums based on these sustainability and decarbonization coefficients and the price coefficients are shown in table 7.

These WTP estimates show the distribution of the marginal WTP for coffee with the given label compared to filter coffee without the given label all other things equal. The interpretation of these numbers is that, on average, buyers would be equally divided between 500g (ground or whole bean) filter coffee without the attribute and an identical package with the additional attribute that costs original price plus the given amount (in €). The WTP premium of group 1 differs from group 2, as the price coefficient of group 2 was adapted based on the interaction term.

A violin plot of the WTP distributions across the full sample including a simplified boxplot is provided in figure 2. The diameter of the violin plots on the y-axis correspond to the relative frequency of individuals in the sample with the given marginal WTP shown on the x-axis scale i.e., the width of the grey area corresponds to the proportion of participants with the given WTP premium. The simplified boxplots provide a visualization of the overall median and quartiles shown in Table 7.

The plot shows that while preferences vary widely, the distributions do not indicate unexpected distribution patterns, such as distinct clusters with opposing preferences. Preference differences like these can be missed if only coefficient estimates and boxplots and not the individual parameter estimates are considered.

4.5.2. Results in Relation to the Hypotheses

In the following paragraphs, the presented results are evaluated in relation to the initial hypotheses one to four introduced in chapter 2.3.

> **H1:** On average, participants are willing to pay a premium for filter coffee that was shipped without causing net emissions.

The result of the random parameter logit shows significant estimates for emission reductions in maritime shipping ($M = 2.61$, $SE = .31$, $p<.001$), emission reductions for the full supply chain ($M = 3,67$, $SE = .35$, $p<.001$), offsets of the shipping emissions ($M = 2.50$, $SE = .21$, $p<.001$), and offsets for emissions of the full supply chain ($M = 2.73$, $SE = .30$, $p<.001$), supporting H1.

The resulting WTP derived from the individuals' level estimates and price coefficients are shown in Table 7 and Figure 2. Participants are indeed willing to pay a premium for filter coffee that was shipped without causing net emissions.

> **H2:** On average, participants value decarbonization of the full supply chain more than decarbonization of parts of the supply chain (maritime shipping only).

A two-tailed z-test showed a significant difference between the coefficients for full and partial direct reductions of emissions ($z = 2.26$, $p = .02$), but no significant difference between the two coefficients for offsets ($z = .57$, $p = .57$).

This implies that participants were valuing decarbonization of the full supply chain more than a decarbonization of the maritime shipping in the context of direct reductions but not in the context of carbon offsets. Thus, the data does not fully support H2. Counter to expectations, participants only valued decarbonization of the supply chain significantly more for direct reductions.

> **H3:** On average, participants attribute more value to direct emissions reductions compared indirect offsets.

The model coefficient for maritime shipping with direct reductions was not significantly higher than the corresponding coefficient for offsets in a one-tailed z-test, $z = .27$, $p = .78$. The coefficient of the supply chain was however significantly higher for full decarbonization compared to offsets in a second one-tailed z-test, $z = 2.05$, $p = .02$. H3 can thus not be fully supported. While participants attributed significantly less value to offsets compared to direct emission reductions for the full supply chain, they did not do so for maritime shipping.

> **H4:** On average, participants that prefer ground or whole bean filter coffee have identical preferences for decarbonization, sustainability labels and price.

The hypothesis could not be fully supported. While participants from the groups did indeed not differ significantly in their underlying preferences for decarbonization and sustainability labels, the group whole bean coffee differed significantly from group ground bean in terms of their price coefficient as shown in the significant interaction term between group and price coefficient ($M = .62$ $SE = .09$, $p<.001$). Group 1 was thus willing to pay more for coffee in general which effectively resulted in in higher WTP for sustainability labels on the package.

5. Discussion

The presented data results in a coherent description of participant behavior and stated preferences. Stated importance ratings were aligned with the behavioral measurement during the DCE. Fairtrade was both stated as one of the most important criteria when buying coffee and resulted in the biggest WTP increase. Full reductions of emissions across the supply chain were valued at similar importance as Fairtrade

Table 7: Marginal WTP for Sustainability and Decarbonization ($N = 299$)

	Full Sample			Group 1: Ground bean			Group 2: Whole bean		
Parameter	**Q1**	***Mdn***	**Q3**	**Q1**	***Mdn***	**Q3**	**Q1**	***Mdn***	**Q3**
Fairtrade label	0.94	2.77	5.62	1.04	2.27	4.29	0.81	3.32	5.94
Organic label	0.58	1.61	3.30	0.50	1.48	2.99	0.65	1.81	3.86
Offset of maritime emissions	0.46	1.79	3.88	0.65	1.78	3.19	0.43	1.92	4.29
Offset of full supply chain	0.47	1.89	4.25	0.47	1.80	3.69	0.46	2.05	4.64
Reduction of maritime emissions	0.63	1.95	3.95	0.55	1.90	3.38	0.64	2.11	4.90
Reduction of full supply chain	1.37	2.82	5.22	1.21	2.48	4.62	1.46	3.08	6.02

Note: estimates in € ; Mdn = Median, Q1 and Q3 = 1st and 3rd distribution quartiles.

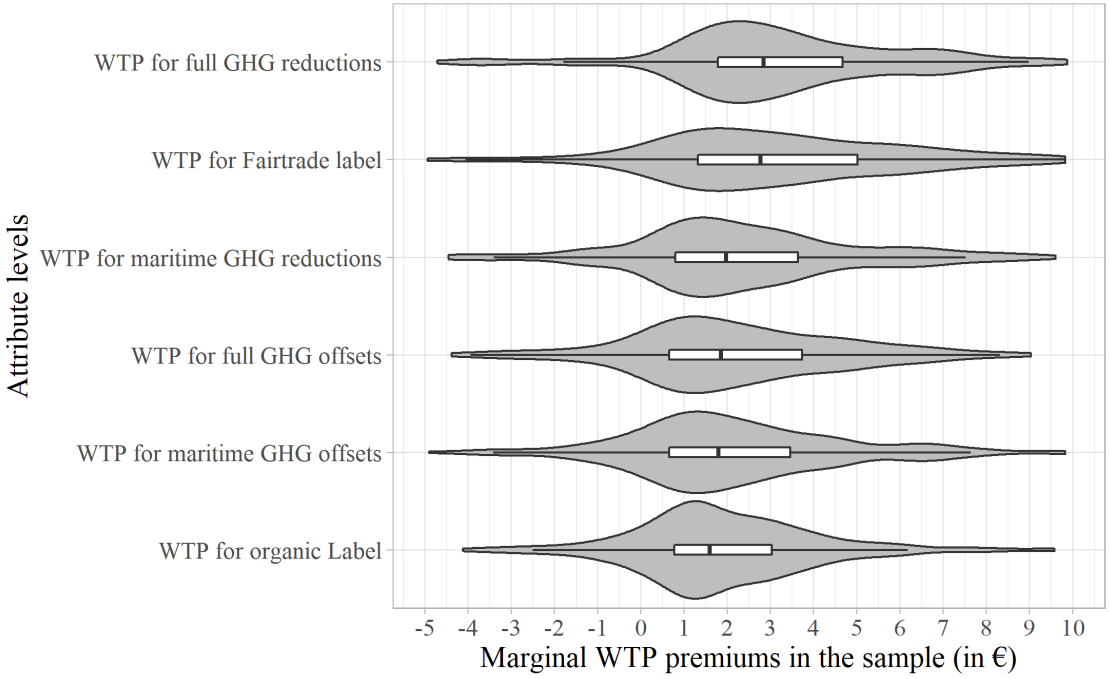

Figure 2: Distribution of Marginal WTP Premium in the Sample ($N = 299$)

certification in the DCE. The three other decarbonization labels influenced decisions less than the Fairtrade certification and full supply chain emission reductions. Nonetheless, all three decarbonization labels exerted a significant, moderate influence on decisions made. The slightly lesser valued organic label also exerted a significant positive effect on the choice. These DCE results are aligned with the stated preference ratings indicating a similar order of priority, except for the label "full decarbonization of the supply chain" which performed even better than expected in comparison to the other labels.

Stated importance of price was ranked at medium importance with a standard deviation of approximately one rank. This heterogeneity was also present in the price coefficient and is not unexpected as the in-store price comparison revealed widely varying market prices for coffee based on brands, labels, and quality. The small demographic and stated preference differences between group 1 and group 2 did not affect their modelled underlying preference for sustainability labels. However, group 2 did state a significantly lower importance of price (weak effect), an effect that was also measured and modelled in the DCE. This results in a higher WTP for coffee in general and thus effectively in higher WTP for sustainability labels on the package. While group 2 was effectively indeed willing to pay more for sustainability and decarbonization labels, they were not necessarily valuing the sustainability and decarbonization labels more compared to group 1, but instead cared less about prices for coffee in general. The average underlying value derived from sustainability labels and decarbonization seems to be rather stable across these segments. Analyzing the stated frequently bought coffees of the groups resulted in slightly

conflicting findings and did not show significant differences in reported price per kg paid between groups. However, these results relied on remembered average prices of small subsamples, thus these subtle differences revealed in the DCE might be noticeable in this less sensitive measurement.

Group 2 was more particular about their coffee choice and tended to choose the no-choice option significantly more often than group 1. Subtle group differences that might explain these results include that whole bean users were on multiple dimensions slightly more interested in product quality criteria, such as taste, smell, type of roast, and digestibility. On average, group 2 also stated to credit slightly lower importance to convenience, coherent with the fact that on average they must obviously put more effort into the preparation of their coffee. While a slightly higher age and a higher rate of students was found in group 1, there was no gender difference between groups. These results provide coherent insights into the sample structure and the stated importance ratings, but the validity of any single significant associations should not be overestimated. Most associations are highly unlikely to be random, but as the beta error was not controlled for in this context, some of the results can be random given the size of the underlying correlation matrix. No significant group differences in the importance of sustainability dimensions (e.g., organic and Fairtrade labelling) were found in the DCE.

Counter to expectations, hypotheses one and two concerned with differences in the valuation of different degrees of decarbonization could not be fully supported. The individual parameter distributions, if ordered by Median, resemble the order of valuation that was expected by the hypothesis, full decarbonization performed better than partial decarbonization and reductions performed better than offsets. These differences were however not big enough to lead to a statistically significant difference between these coefficients for full and partial offsets and for maritime shipping reductions and offsets. This could be based either on these differences being too small for the sample and design to detect or might be the result of actual indecisiveness of the consumers on these topics. Future research including stated preference data and a stronger focus on comprehension of the various forms of offsets and decarbonization could reveal possible reasons for indecisiveness. The comprehension measurements indicate that participants were worse at assigning the situational descriptions to the correct terms for situations describing offsets compared to situations describing reductions. The low overall score average is an indication for a lack of comprehension of terminology and contents and might be one explanation for the similar valuation of full and partial offsets. There were no comprehension performance differences between the two groups.

Price coefficient and coefficients for the labels resulted in plausible WTP ranges when comparing the empirical findings of the choice experiment to the prices for certified coffees as reported in the preliminary in-store price comparison. The results indicate a different priority order of WTP when comparing the results to the work of van Loo et al. (2015). While the organic label WTP premium is estimated very similarly, Fairtrade was less important to the American sample and the carbon footprint reduction pledge presented did not influence decision making significantly. Given the different content and context of the carbon label used by van Loo et al. (2015), the different and higher valuation of the carbon offset and reduction labels presented in this study seems viable. The authors did not provide a description of the label's contents beyond the vague pledge "reducing our carbon emissions", while this study's labels indicated more concrete and tangible emission reductions. Furthermore, it is not unlikely that a 2015 American sample will behave differently than a 2021 German sample in similar context. The valuation of various sustainability aspects is expected to differ across time, culture, and demographics and might have increased from 2015 to 2021 given the rise of sustainability trends.

Comparing the results to other relevant studies, at first glance the WTP premium of the whole bean segment for offsetting maritime shipping ($M = 1.92$€ per 500g) and the full supply chain ($M = 2.05$€ per 500g) look very similar to the WTP for a "carbon neutral label" reported by Birkenberg et al. (2021) if the packaging size is not considered ($M = 1.77$€ per 250g of Arabica whole beans). Fairtrade was valued less in the presented sample ($M = 3.32$€ per 500g) compared to the estimates of Birkenberg et al. ($M = 4.30$€ per 250g). These deviations might be the result of different design and sampling methods. Amongst other differences, their results rely on a sample recruited personally in German coffee shops that was informed about GHG emissions of the coffee life cycle before the experiment, possibly influencing WTP for carbon net neutrality. Other differences include the survey in a shop setting possibly priming for more expensive premium or specialty coffee of a specific brand. The public setting might also be more prone to subtle influences of social desirability compared to the online setting of this study. Furthermore, their overall sample was smaller, and the resulting model less precise in terms of standard errors compared to the model introduced. Future DCEs might consider adding varying packaging sizes as an attribute. Interestingly although possibly randomly, if the packaging size is not considered WTP for the sustainable labels seems aligned in both models. If differences in design and setting do not explain the deviations, WTP for coffee might in customer perception be less of a relative value dependent on packaging size and more of an absolute value per package of coffee bought. This could make premium segments that rely on smaller bags of coffee an attractive initial niche market for carbon reduction. Rotaris and Danielis (2011) reported a high premium for Fairtrade coffee ($M = 2.20$€ per 250g package of beans or ground coffee) in an Italian sample ($N = 135$), while Lappeman et al. (2019) reported a lower value ($M = \$1.22$ per 250g) in a South African sample ($N = 300$). Considering these international deviations, the Fairtrade WTP modelled in this study (which can be an indicator of reliability of the other labels) appears to be in a plausible range. The results of Schniederjans and Starkey (2014) and Polinori et al. (2018)

are hard to relate to this study given their different use case, design, and label descriptions leaving room for subjective interpretation of "green transportation".

Heterogeneity of WTP for the labels is evident in the significant *SD* estimates of the random parameters and visualized in Figure 2. Plotting the individual WTP premiums on a violin graph illustrated that the marginal WTP premium of most individuals was (often strongly) increased by the labels, but some individuals were also discouraged by them. This result was expected based on the findings of previous authors (Birkenberg et al., 2021; Lappeman et al., 2019; Rotaris & Danielis, 2011; van Loo et al., 2015) and is coherent with the heterogenous importance ratings and reported prices of favorite coffee. A business implication of these individual differences is possible potential for market skimming when first introducing GHG reductions (Simon & Fassnacht, 2016). Revenues generated by initially realizing higher prices with labelled coffee, specifically targeting niche segments with higher WTP, could be a vital driver to fund the necessary growth to eventually compete with container shipping at scale. Identifying and describing resulting segments of interest lies beyond the scope of this study, nonetheless whole bean coffee consumers might be a first segment to consider because of their increased overall WTP. Furthermore, the correlations between the decarbonization and sustainability labels shows interest in existing sustainability labels is related to interest in decarbonization. Whole bean customers preferring organic and Fairtrade coffee could thus be an initial audience to target.

6. Limitations

The presented study provides an applied example for pricing sustainable shipping on the use case of coffee. Multiple precautions were taken to ensure internal validity of the results, including pre-tests of the final survey, prior research to design the experiment in a suitable manner (e.g., defining meaningful price levels based on store data), consistent randomization to avoid maturation and testing effects, and limited temporary survey availability to avoid confounding effects through external events. Information exchange between participants can be assumed as minimal in the chosen online setting. The anonymous online setting minimizes investigator influence and all wordings were carefully chosen to avoid priming and unintended bias. The questionnaire is provided in Appendix A for reference and transparency. The proposed methodology closely mirrors common choice behavior in purchasing settings, however, proper framing of the proposed emission labels and a real buying context (including brands and more product variety) might moderate the observed effects and strengthen or weaken individuals' WTP.

Possible selection biases cannot be fully mitigated as a convenience sample was used. While the sample is transparently defined and rather balanced it shows tendencies towards higher education, females and individuals living in southern Germany. Future research based on representative samples is needed to establish findings that are fully representative of the German market. Measuring the influence of income and/or wealth on the price coefficient can be an interesting addition for future inquiry. While incentives could have been an additional measure to reach a bigger sample size, the statistical efficiency of the proposed modelling process led to a coherent collection of significant findings. The dropout rate of 31% might seem high for a study of the given length of around 10 minutes. A high dropout rate was however expected, as no incentives for completion were provided and the 14 consecutive choice tasks in the middle of the survey were demanding on participants. Most of the lost participants dropped out during these tasks. When estimating a viable sample in reach at not more than ∼300 participants, the choice was made to rather include more choice sets across fewer participants than to introduce fewer choice sets with possibly higher completion rates. Eventually more participants were reached, but a significant dropout and possible resulting selection effects were accepted.

The external validity of the results, i.e., whether they can be generalized, is another quality criterion to consider. Of course, the example use case of coffee cannot provide a context-independent valid estimate of WTP for sustainable (maritime) shipping, but it does provide an exemplary estimate in a rigorously controlled setting of interest. The research was not based on a representative sample of the German population, it does however provide an estimate of the size of the effects of inquiry, that needs to be measured, validated, and adapted in other contexts.

Concerns that could confound the results were monitored and thoroughly analyzed for this study. As a conclusion from the qualitative analysis of these concerns, while some concerns did arise, none threaten the overall validity of the results. Two possible factors of influence were revealed in the auxiliary parts of the survey: lacking knowledge about or comprehension of the concepts offsetting and emission reductions as well as credibility concerns regarding labels. In retrospect, an additional measure would have been useful to assess the concurrent validity of the results of the comprehension scale indicating a moderate to low average knowledge of the detailed meaning of the concepts. The categories revealed in the qualitative analysis of concerns should be considered in further research on similar labels. Future labels need to be optimized for effectiveness and easy comprehension, transparency and tangible clarity of certification. Distrust in labels, greenwashing concerns and possible doubts on the technical feasibility of emission reduction measures need to be considered and handled in this context.

Low knowledge levels of offsets and reductions were measured but could be a result of the scale used. While great care was taken in defining the scale and examples used in line with the definitions of the terms, the scale has not been validated and tested for reliability. Anyhow, low comprehension levels are not unlikely given the evidence from previous international research (Polonsky et al., 2011, 2015; Tao et al., 2021). No notable comprehension differences between the two groups were found making it unlikely that these

differences could confound results.

From a modelling perspective, outcomes of advanced statistical models can differ depending on the parameters chosen and the variables considered in the modelling process, however all criteria used for fitting and choosing the models were both transparently provided and discussed in the data analysis section (chapter 3.5) in detail to warrant objectivity and reliability of the research.

7. Conclusion

The goal of this study was to price sustainable i.e., low GHG emitting maritime shipping. Employing coffee as an example, significant willingness to pay premiums for sustainable shipping were found in the sample for fully decarbonized sustainable supply chains, but also for partial and full carbon offsets. Participants were willing to pay significantly more for full technological reductions of the supply chain emissions than for reductions of shipping emissions only and partial and full emission offsets, the latter three resulting in similar WTP. Full decarbonization of the supply chain and Fairtrade certification were associated with the highest willingness to pay premiums in the experiment. Organic certification was also perceived as a benefit and resulted in a significant, albeit smaller WTP premium. An in-store price comparison, as well as self-reported importance ratings and auxiliary survey data supported and complemented these findings.

Participants that prefer whole bean coffee to ground beans did not differ in their underlying preferences for the labels. They did however differ significantly on their price preferences and effectively *displayed a higher overall WTP for all attributes including the sustainability labels*. These differences were coherent with the group's self-reported importance ratings. Whole bean customers claimed to place more importance on quality criteria (e.g., smell and taste) and less importance on price.

Significant heterogeneity was present across all labels and groups, emphasizing the importance of segmentation and targeting for industry practitioners, as well as the importance of further inquiry into the antecedents of this heterogeneity for researchers. The benefits of employing random parameter logit models that can cope with the heterogeneity in this domain of research became evident in the modelling process. The antecedents of this heterogeneity were beyond the scope of this work, however based on existing literature known contextual and personal factors influencing decisions and attitudes in the context were provided.

The presented results extend the domain of research by providing a rigorous measurement of customer preferences and WTP for sustainable shipping and emission reductions across the supply chain. To the best of the author's knowledge, the presented discrete choice experiment is the first to assess the perceived customer value of offsets compared to direct reductions in a controlled setting for a common product use case. The study further provides an applied example for designing and performing an advanced discrete choice experiment, solely relying on openly available open-source software. The practical value of this work is to provide CargoKite and other companies and initiatives with a focus on decarbonization with reliable baseline data that shows a customer demand for carbon neutral (maritime) cargo transport.

Future research is needed to understand the underlying antecedents of the WTP heterogeneity revealed in the sample and the existing literature and validate the robustness of the results in other contexts. As significant differences across international samples were identified in literature, another interesting avenue of research could be to relate the topic to cultural dimensions and societal trends. While exceeding the scope of this study, employing segmentation approaches (e.g., cluster analysis) can be useful to better understand the individual differences that were modelled in the random parameter approach and yield useful results for stakeholders in the field of sustainable maritime shipping, e.g., by identifying and describing segments of interest and increased WTP. Furthermore, while existing research provides some interesting and actionable insights into the effective design of offset programs, labelling and related issues, industry practitioners could benefit from further research in this area.

References

Akaichi, F., de Grauw, S., Darmon, P., & Revoredo-Giha, C. (2016). Does Fair Trade Compete with Carbon Footprint and Organic Attributes in the Eyes of Consumers? Results from a Pilot Study in Scotland, The Netherlands and France. *Journal of Agricultural and Environmental Ethics*, *29*(6), 969–984. doi: 10.1007/s10806-016-9642-7

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Iannone, R. (2021). *rmarkdown: Dynamic Documents for R.* Retrieved from https://bookdown.org/yihui/rmarkdown/

Baranzini, A., Borzykowski, N., & Carattini, S. (2018). Carbon offsets out of the woods? Acceptability of domestic vs. international reforestation programmes in the lab. *Journal of Forest Economics*, *32*, 1–12. doi: 10.1016/j.jfe.2018.02.004

Berger, J. (2019). Signaling can increase consumers' willingness to pay for green products. Theoretical model and experimental evidence. *Journal of Consumer Behaviour*, *18*(3), 233–246. doi: 10.1002/cb.1760

Birkenberg, A., Narjes, M. E., Weinmann, B., & Birner, R. (2021). The potential of carbon neutral labeling to engage coffee consumers in climate change mitigation. *Journal of Cleaner Production*, *278*, 123621. doi: 10.1016/j.jclepro.2020.123621

Blasch, J., & Ohndorf, M. (2015). Altruism, moral norms and social approval: Joint determinants of individual offset behavior. *Ecological Economics*, *116*, 251–260. doi: 10.1016/j.ecolecon.2015.04.024

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. doi: 10.1016/j.trb.2009.12.004

Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P.-F. (2006). The No-Choice Option and Dual Response Choice Designs. *Marketing Letters*, *17*(4), 255–268. Retrieved from http://www.jstor.org/stable/40216693

Bridges, J. F. P., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., … Mauskopf, J. (2011). Conjoint analysis applications in health–a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, *14*(4), 403–413. Retrieved 15/07/2021, from https://www.valueinhealthjournal.com/article/S1098-3015(10)00083-5/pdf doi: 10.1016/j.jval.2010.11.013

Canzi, G., Clough, K., & Kronick, C. (2006). Joint statement on offsetting carbon emissions. *Friends of the Earth, Greenpeace and WWF-UK*. Retrieved 07/07/2021, from http://assets.wwf.org.uk/downloads/august06.pdf

Caputo, V., Nayga, R. M., & Scarpa, R. (2013). Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Australian Journal of Agricultural and Resource Economics*, *57*(4), 465–482. doi: 10.1111/1467-8489.12014

Carson, R. T., Flores, N. E., Martin, K. M., & Wright, J. L. (1996). Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods. *Land Economics*, *72*(1), 80. doi: 10.2307/3147159

Carson, R. T., & Louviere, J. J. (2011). A Common Nomenclature for Stated Preference Elicitation Approaches. *Environmental and Resource Economics*, *49*(4), 539–559. doi: 10.1007/s10640-010-9450-x

Chapman, C., & Feit, E. M. (2019). *R For Marketing Research and Analytics*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-14316-9

Chi, O. H., Denton, G., & Gursoy, D. (2021). Interactive effects of message framing and information content on carbon offsetting behaviors. *Tourism Management*, *83*, 104244. doi: 10.1016/j.tourman.2020.104244

Chongsuvivatwong, V. (2018). *epiDisplay: Epidemiological Data Display Package.* Retrieved from https://CRAN.R-project.org/package=epiDisplay

Corbin, J. M., & Strauss, A. L. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory / Juliet Corbin, International Institute for Qualitative Methodology, Anselm Strauss* (Fourth edition ed.). Los Angeles: SAGE.

Creutzig, F., et al. (Eds.). (2014). *Transport*. Cambridge: Cambridge University Press. Retrieved 05.07.2021, from https://www.ipcc.ch/site/assets/uploads/2018/02/ipcc_wg3_ar5_chapter8.pdf doi: 10.1017/CBO9781107415416

Croissant, Y. (2020). Estimation of Random Utility Models in R : The mlogit Package. *Journal of Statistical Software*, *95*(11). doi: 10.18637/jss.v095.i11

Croissant, Y. (2021). *dfidx: Indexed Data Frames.* Retrieved from https://CRAN.R-project.org/package=dfidx

Day, R. (2021). Carbon Offsets Are About To Become A Huge Market. *Forbes*. Retrieved 06/07/2021, from https://www.forbes.com/sites/robday/2021/04/29/carbon-offsets-are-about-to-become-a-huge-market/?sh=6e11971461ec

de Jong, M. D. T., Harkink, K. M., & Barth, S. (2018). Making Green Stuff? Effects of Corporate Greenwashing on Consumers. *Journal of business and technical communication*, *32*(1), 77–112. doi: 10.1177/1050651917729863

Deutscher Kaffeemarkt e.V. (2021). *Deutscher Kaffeemarkt 2020.* Hamburg. Retrieved 21/07/2021, from https://www.kaffeeverband.de/de/presse/deutscher-kaffeemarkt-2020-20-tassen-pro-kopf-mehr-konsumiert#

Deutscher Kaffeeverband e.V. . (2020). *Handel, Kaffeepreis und Logistik: Transport des Rohkaffees.* Hamburg. Retrieved 04.07.2021, from https://www.kaffeeverband.de/de/kaffeewissen/handel

Drichoutis, A. C., Lusk, J. L., & Pappa, V. (2016). Elicitation formats and the WTA/WTP gap: A study of climate neutral foods. *Food Policy*, *61*, 141–155. doi: 10.1016/j.foodpol.2016.03.001

EcoClipper. (2021). *EcoClipper | Emission Free Shipping and Travel.* Alkmaar, Netherlands. Retrieved 15/07/2021, from https://ecoclipper.org/

Egloff, C., Escudero, R. G., Sanders, U., Webster, J., & Zampelas, A. (2019). Sharing the Costs of IMO 2020 Across the Ecosystem. *Germany - DE*. Retrieved 27/08/2021, from https://www.bcg.com/de-de/publications/2019/sharing-costs-imo-2020-across-ecosystem

European Commission. (2021). *The organic logo: Aims of the organic logo.* Retrieved 06/07/2021, from https://ec.europa.eu/info/food-farming-fisheries/farming/organic-farming/organic-logo_de

Faber, J., & Kleijn, A. (2020). Fourth IMO Greenhouse Gas Study: Final Report. *CE Delft*. Retrieved 11.01.2020, from https://www.cedelft.eu/en/publications/2488/fourth-imo-greenhouse-gas-study

Fairtrade International. (2021a). *Absatzentwicklung und Details über Fairtrade-Kaffee in Deutschland.* Bonn, Germany. Retrieved 15/09/2021, from https://www.fairtrade-deutschland.de/produkte/kaffee/hintergrund-fairtrade-kaffee.html

Fairtrade International. (2021b). *Aims of the Fairtrade Standards.* Bonn, Germany. Retrieved 06/07/2021, from https://www.fairtrade.net/standard/aims

Felser, G. (2015). *Werbe- und Konsumentenpsychologie*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-37645-0

Fiedler, M., Frank, K., & Volland, M. (2020). *Aktuelle Entwicklungen im Fairen Handel: Umsatzentwicklungen im Geschäftsjahr 2019.* Berlin. Retrieved 04.07.2021, from https://www.forum-fairer-handel.de/fileadmin/user_upload/dateien/publikationen/materialien_des_ffh/2020_Aktuelle-Entwicklungen_im_FH_web.pdf

Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll / Andy Field* (4th ed. ed.). London: SAGE.

Förster, H. (2020). *Verbrauchs- und Medienanalyse - VuMA 2021.* Retrieved 04.07.2021, from https://www.vuma.de/fileadmin/user_upload/PDF/berichtsbaende/VuMA_Berichtsband_2021.pdf

Galik, C. S., & Jackson, R. B. (2009). Risks to forest carbon offset projects in a changing climate. *Forest Ecology and Management*, *257*(11), 2209–2216. doi: 10.1016/j.foreco.2009.03.017

Gustafsson, A., Herrmann, A., & Huber, F. (2007). *Conjoint Measurement: Methods and Applications* (4th ed. ed.). Berlin: Springer.

Harrell, F. E., Jr. (2021). *Hmisc: Harrell Miscellaneous.* Retrieved from https://CRAN.R-project.org/package=Hmisc

Hasselbach, J. L., & Roosen, J. (2015). Motivations behind Preferences for Local or Organic Food. *Journal of International Consumer Marketing*, *27*(4), 295–306. doi: 10.1080/08961530.2015.1022921

Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G. M., Prior, T., Marshall, D. A., Cunningham, C., … Bridges, J. F. P. (2016). Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, *19*(4), 300–315. doi: 10.1016/j.jval.2016.04.004

Hensher, D. A., Rose, J. M., & Greene, W. H. (2015). *Applied choice analysis* (2nd edition ed.). Cambridge: Cambridge University Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy1503/2014043411-b.html

Herrmann, S. (2015). *Bewertung der Lebensmittel-Labels 2015: Hintergrundbericht.* Zürich. Retrieved 15/09/2021, from https://www.wwf.ch/sites/default/files/doc-2017-10/2015-11-Hintergrundbericht-Lebensmittellabel-de.pdf

Hess, S., & Daly, A. J. (2014). *Handbook of choice modelling*. Cheltenham: Edward Elgar.

Hole, A. R. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health economics*, *16*(8), 827–840. doi: 10.1002/hec.1197

Huber, R. A., Anderson, B., & Bernauer, T. (2018). Can social norm interventions promote voluntary pro environmental action? *Environmental Science & Policy*, *89*, 231–246. doi: 10.1016/j.envsci.2018.07.016

International Coffee Organization. (2021). *World coffee consumption.* Retrieved 07/06/2021, from http://www.ico.org/prices/new-consumption-table.pdf

International Maritime Organization. (2020a). *IMO Environment Committee approves amendments to cut ship emissions.* London. Retrieved 06.07.2021, from https://www.imo.org/en/MediaCentre/PressBriefings/pages/42-MEPC-short-term-measure.aspx

International Maritime Organization. (2020b). *IMO working group agrees further measures to cut ship emissions.* London. Retrieved 06.07.2021, from https://www.imo.org/en/MediaCentre/PressBriefings/pages/36-ISWG-GHG-7.aspx

International Maritime Organization. (2020c). *Reducing greenhouse gas emissions from ships.* London. Retrieved 19.01.2021, from https://www.imo.org/en/MediaCentre/HotTopics/Pages/Reducing-greenhouse-gas-emissions-from-ships.aspx

Johnson, R., & Orme, B. (2003). Getting the most from CBC. *Sequim: Sawtooth Software Research Paper Series, Sawtooth Software*.

Kamakura, W., Haaijer, R., & Wedel, M. (2001). The No—Choice Alternative in Conjoint Choice Experiments. *International Journal of Market Research*, *43*(1), 1–12. doi: 10.1177/147078530104300105

Klein, J. (2021). *Unsere Klimaziele: Fairer Handel mit Kaffeebohnen, fair, bio und klimaneutral: Transport per Frachtsegler.* Leipzig. Retrieved 15/07/2021, from https://cafe-chavalo.de/content/17-unsere-klimaziele

Kollmuss, A. (2010). *Handbook of carbon offset programs: Trading systems, funds, protocols and standards / Anja Kollmuss ... [et al.].* London: Earthscan.

Lappeman, J., Orpwood, T., Russell, M., Zeller, T., & Jansson, J. (2019). Personal values and willingness to pay for fair trade coffee in Cape Town, South Africa. *Journal of Cleaner Production*, *239*, 118012. doi: 10.1016/j.jclepro.2019.118012

Leifeld, P. (2013). texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables. , *55*(8), 1–24. Retrieved from http://dx.doi.org/10.18637/jss.v055.i08

Lera-López, F., Faulin, J., & Sánchez, M. (2012). Determinants of the willingness-to-pay for reducing the environmental impacts of road transportation. *Transportation Research Part D: Transport and Environment*, *17*(3), 215–220. doi: 10.1016/j.trd.2011.11.002

List, J. A., & Gallet, C. A. (2001). What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? *Environmental and Resource Economics*, *20*(3), 241–254. doi: 10.1023/A:1012791822804

Liu, M., & Wronski, L. (2018). Examining Completion Rates in Web Surveys via Over 25,000 Real-World Surveys. *Social Science Computer Review*, *36*(1), 116–124. doi: 10.1177/0894439317695581

Loomis, J. B. (2014). 2013 WAEA Keynote Address: Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys. *Journal of Agricultural and Resource Economics*, *39*(1), 34–46. Retrieved from http://www.jstor.org/stable/44131313

Lusk, J. L. (2003). Effects of Cheap Talk on Consumer Willingness–to–Pay for Golden Rice. *American Journal of Agricultural Economics*, *85*(4), 840–856. doi: 10.1111/1467-8276.00492

Meunier, C., Stoll, J., & Schoen, L. (2019). *Freiwillige CO2-Kompensation.* Dessau-Roßlau. Retrieved 07/07/2021, from https://www.umweltbundesamt.de/themen/freiwillige-co2-kompensation

Murphy, J. J., Allen, P. G., Stevens, T. H., & Weatherhead, D. (2005). A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics*, *30*(3), 313–325. doi: 10.1007/s10640-004-3332-z

Norwood, F. B., & Lusk, J. L. (2011). Social Desirability Bias in Real, Hypothetical, and Inferred Valuation Experiments. *American Journal of Agricultural Economics*, *93*(2), 528–534. doi: 10.1093/ajae/aaq142

Oxford University Press. (2021). *Greenwashing - Definition.* Oxford. Retrieved 26/07/2021, from https://www.oxfordlearnersdictionaries.com/definition/english/greenwash

Parker, J. R., & Schrift, R. Y. (2011). Rejectable Choice Sets: How Seemingly Irrelevant No-Choice Options Affect Consumer Decision Processes. *Journal of Marketing Research*, *48*(5), 840–854. doi: 10.1509/jmkr.48.5.840

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*, *36*(4), 859–866. doi: 10.1111/j.1745-9125.1998.tb01268.x

Pitt, A. (2017). *IMO 2020 Regulation Could Cost Shippers Extra US$60 Billion a year.* Singapore. Retrieved 14.07.2021, from https://www.woodmac.com/press-releases/imo-2020-regulation-shippers-us$60-billionyear/

Pleeging, E., van Exel, J., Burger, M. J., & Stavropoulos, S. (2021). Hope for the future and willingness to pay for sustainable energy. *Ecological Economics*, *181*, 106900. doi: 10.1016/j.ecolecon.2020.106900

Polinori, P., Marcucci, E., Gatta, V., Bigerna, S., Bollino, C. A., & Micheli, S. (2018). Eco-Labeling And Sustainable Urban Freight Transport: How Much Are People Willing To Pay For Green Logistics ? *Rivista Internazionale di Economia dei Transporti / International Journal of Transport Economics*, *XLV*, 631–658. doi: 10.19272/201806704006

Polonsky, M. J., Garma, R., & Landreth Grau, S. (2011). Western consumers' understanding of carbon offsets and its relationship to behavior. *Asia Pacific Journal of Marketing and Logistics*, *23*(5), 583–603. doi: 10.1108/13555851111183048

Polonsky, M. J., Grau, S. L., & Garma, R. (2015). Exploring US Consumers Understanding of Carbon Offsets. In L. Robinson (Ed.), *Proceedings of the 2009 Academy of Marketing Science (AMS) Annual Conference* (pp. 14–18). Cham: Springer International Publishing. doi: 10.1007/978-3-319-10864-3{\textunderscore}10

Psaraftis, H. N. (2019). *Sustainable Shipping*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-04330-8

Qualtrics. (2021). *Qualtrics XM.* Provo, Utah, USA: Author. Retrieved 17.02.2021, from https://www.qualtrics.com

R Core Team. (2021). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rao, V. R. (2014). *Applied Conjoint Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-87753-0

Raza, Z. (2020). Effects of regulation-driven green innovations on short sea shipping's environmental and economic performance. *Transportation Research Part D: Transport and Environment*, *84*, 102340. doi: 10.1016/j.trd.2020.102340

Ritchie, B. W., Kemperman, A., & Dolnicar, S. (2021). Which types of product attributes lead to aviation voluntary carbon offsetting among air passengers? *Tourism Management*, *85*, 104276. doi: 10.1016/j.tourman.2020.104276

Rose, J. M., & Bliemer, M. C. J. (2013). Sample size requirements for stated choice experiments. *Transportation*, *40*(5), 1021–1041. doi: 10.1007/s11116-013-9451-z

Rotaris, L., & Danielis, R. (2011). Willingness to Pay for Fair Trade Coffee: A Conjoint Analysis Experiment with Italian Consumers. *Journal of Agricultural & Food Industrial Organization*, *9*(1). doi: 10.2202/1542-0485.1305

RStudio Team. (2021). *RStudio: Integrated Development Environment for R.* Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/

Schmidt, J., & Bijmolt, T. H. A. (2020). Accurately measuring willingness to pay for consumer goods: a meta-analysis of the hypothetical bias. *Journal of the Academy of Marketing Science*, *48*(3), 499–518. doi: 10.1007/s11747-019-00666-6

Schniederjans, D. G., & Starkey, C. M. (2014). Intention and willingness to pay for green freight transportation: An empirical examination. *Transportation Research Part D: Transport and Environment*, *31*, 116–125. doi: 10.1016/j.trd.2014.05.024

Schwirplies, C., Dütschke, E., Schleich, J., & Ziegler, A. (2017). *Consumers' willingness to offset their CO2 emissions from traveling: A discrete choice analysis of framing and provider contributions: Working Papers "Sustainability and Innovation* (No. S05/2017). Retrieved from https://ideas.repec.org/p/zbw/fisisi/s052017.html

Schwirplies, C., Dütschke, E., Schleich, J., & Ziegler, A. (2019). The willingness to offset CO2 emissions from traveling: Findings from discrete choice experiments with different framings. *Ecological Economics*, *165*, 106384. doi: 10.1016/j.ecolecon.2019.106384

Serra, P., & Fancello, G. (2020). Towards the IMO's GHG Goals: A Critical Overview of the Perspectives and Challenges of the Main Options for Decarbonizing International Shipping. *Sustainability*, *12*(8), 3220. doi: 10.3390/su12083220

Simon, H., & Fassnacht, M. (2016). *Preismanagement*. Wiesbaden: Springer Fachmedien Wiesbaden. doi: 10.1007/978-3-658-11871-6

Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 112–141. doi: 10.1177/0002716212458082

Sonnenschein, J., & Smedby, N. (2019). Designing air ticket taxes for climate change mitigation: insights from a Swedish valuation study. *Climate Policy*, *19*(5), 651–663. doi: 10.1080/14693062.2018.1547678

Steiner, B. E., Peschel, A. O., & Grebitus, C. (2017). Multi-Product Category Choices Labeled for Ecological Footprints: Exploring Psychographics and Evolved Psychological Biases for Characterizing Latent Consumer Classes. *Ecological Economics*, *140*, 251–264. doi: 10.1016/j.ecolecon.2017.05.009

Streimikiene, Balezentis, Alisauskaite-Seskiene, Stankuniene, & Simanaviciene. (2019). A Review of Willingness to Pay Studies for Climate Change Mitigation in the Energy Sector. *Energies*, *12*(8), 1481. doi: 10.3390/en12081481

Tao, Y., Duan, M., & Deng, Z. (2021). Using an extended theory of planned behaviour to explain willingness towards voluntary carbon offsetting among Chinese consumers. *Ecological Economics*, *185*, 107068. doi: 10.1016/j.ecolecon.2021.107068

Traets, F., Sanchez, D. G., & Vandebroek, M. (2020). Generating Optimal Designs for Discrete Choice Experiments in R : The idefix Package. *Journal of Statistical Software*, *96*(3). doi: 10.18637/jss.v096.i03

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (Second Edition ed.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511805271

United Nations. (2016). *Maritime Transport Is 'Backbone of Global Trade and the Global Economy', Says Secretary-General Ban Ki-moon in Message for International Day*. London. Retrieved 06.07.2021, from https://www.un.org/press/en/2016/sgsm18129.doc.htm

United Nations Treaty Collection. (2015). *The Paris Agreement*. United Nations Framework Convention on Climate Change. Retrieved 05.07.2021, from https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

Upham, P., Dendler, L., & Bleda, M. (2011). Carbon labelling of grocery products: public perceptions and potential emissions reductions. *Journal of Cleaner Production*, *19*(4), 348–355. doi: 10.1016/j.jclepro.2010.05.014

Usva, K., Sinkko, T., Silvenius, F., Riipi, I., & Heusala, H. (2020). Carbon and water footprint of coffee consumed in Finland—life cycle assessment. *The International Journal of Life Cycle Assessment*, *25*(10), 1976–1990. doi: 10.1007/s11367-020-01799-5

van Loo, E. J., Caputo, V., Nayga, R. M., Seo, H.-S., Zhang, B., & Verbeke, W. (2015). Sustainability labels on coffee: Consumer preferences, willingness-to-pay and visual attention to attributes. *Ecological Economics*, *118*, 215–225. doi: 10.1016/j.ecolecon.2015.07.011

Waring, E., Quinn, M., McNamara, A., Arino de la Rubia, E., Zhu, H., & Ellis, S. (2021). *skimr: Compact and Flexible Summaries of Data.* Retrieved from https://CRAN.R-project.org/package=skimr

Wickham, H. (2021). *tidyr: Tidy Messy Data.* RStudio, PBC. Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Romain, F., Henry, L., & Müller, K. (2021a). *dplyr: A Grammar of Data Manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Romain, F., Henry, L., & Müller, K. (2021b). *ggplot2: Elegant Graphics for Data Analysis.* Retrieved from https://CRAN.R-project.org/package=dplyr

Willer, H., Trávníček, J., Meier, C., & Schlatter, B. (2021). *The World of Organic Agriculture: Statistics & Emerging Trends 2021* (22nd ed.). Bonn. Retrieved 28.09.2021, from https://www.fibl.org/fileadmin/documents/shop/1150-organic-world-2021.pdf

Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R Markdown: The definitive guide*. Boca Raton: Taylor & Francis CRC Press.

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R Markdown cookbook* (1st ed.). Boca Raton: Chapman & Hall/CRC.