Title

Making Better Decisions in Sustainable Operations: Behavioral and Optimization Based Perspectives

Author

Yavuz, Mirel

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Making Better Decisions in Sustainable Operations: Behavioral and Optimization Based Perspectives

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Mirel Yavuz

2024

ABSTRACT OF THE DISSERTATION

Making Better Decisions in Sustainable Operations:
Behavioral and Optimization Based Perspectives

by

Mirel Yavuz
Doctor of Philosophy in Management
University of California, Los Angeles, 2024
Professor Charles J. Corbett, Chair

Decision-makers have access to better environmental information through tools like life-cycle assessment (LCA). However, these methods often implicitly assume that decision-makers make rational assessments when weighing environmental criteria. The impacts of behavioral biases and context effects on decisions remain less recognized, despite being well-documented in behavioral science. Current sustainability approaches offer little guidance in finding optimal solutions when decision-makers' preferences are unknown. Unlike traditional trade-offs involving economic factors, sustainability trade-offs involve intangible and emotionally charged dimensions. Firms aiming for sustainability lack clear guidelines for trading-off environmental and social impacts. This dissertation seeks to assist decision-makers in the context of sustainability from behavioral and optimization-based perspectives.

We first conduct an experiment to test whether decision makers are subjected to two context effects, namely attraction and compromise effects, when faced with environmental trade-offs. Our results show that these context effects are prevalent and substantial in both environmental and non-environmental settings, which highlights the need to integrate

behavioral science into environmental decision-making.

We introduce an interactive optimization method that aims to help decision makers with difficult trade-offs when their value function is not known. This method involves asking decision makers pairwise comparison questions to identify the optimal solution. The approach minimizes the cognitive burden on decision makers by asking fewer, easier questions while still guaranteeing an optimal solution. We test the method in the context of sustainable sourcing in the apparel industry, demonstrating its effectiveness in converging to optimal solutions with fewer decision-making steps compared to traditional methods. Initial feedback from industry practitioners is promising.

An interactive approach like this is uncommon in decision-making in sustainable operations. A more conventional approach would be to elicit the decision-maker's weights and then use a traditional optimization method using those weights. To test which method decision-makers prefer, we also develop an experimental framework, using oTree, to compare the performance of the proposed interactive optimization method with a more traditional approach, eliciting weights through direct rating and then optimizing. Although the experimental framework is initially designed specifically to assess our interactive algorithm, the framework highlights more generally how experiments can be designed that combine Gurobi-based optimization within the experimental environment provided by oTree.

The dissertation of Mirel Yavuz is approved.

Elisa F. Long

Velibor Mišić

Deepak Rajagopal

Auyon A. Siddiq

Charles J. Corbett, Committee Chair

University of California, Los Angeles

2024

To the loving memories of my mother Hülya Aylin Kulan Güneri and my grandmother Şahika Kulan, whom I miss every day...

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Professor Charles Corbett, for his continuous support and guidance over the past years. I have encountered several unfortunate events during my Ph.D. journey, but Charles has always been there for me and helped me overcome every challenge in the best way possible. Charles also taught me how to look at things in different ways, which has made me not just a better researcher, but also a better person overall. Charles: it has been a great pleasure to work with you and learn from you. I cannot fully express how grateful I am to have such a great advisor like you. You have given me all your support in every challenge I faced throughout my Ph.D. My mother always wanted me to forward her words to thank you for caring about my well being and your continuous support. She wanted to say these, in probably a much better way, in person; unfortunately her time in this world did not allow it. So this is not only coming from me, but also coming from her and my grandmother: Thank you a trillion times!

I would like to thank Professors Auyon Siddiq, Deepak Rajagopal, Elisa Long and Velibor Mišić, for being my thesis committee members, giving me critical feedback on my research, and supporting me in the academic job market. Auyon: thank you for helping me both in the early stages of my job market paper as well as later to make my research accessible to a broader audience. Deepak: thank you for expanding my knowledge in sustainability. The conversations we had, especially while working on the context effects paper, were extremely helpful. Elisa: thank you for encouraging me to pursue Ph.D. at UCLA Anderson and all your support, even before I accepted the offer and started the program. Thank you for creating a safe space for me, I knew I could always ask for your advice and feedback. Velibor: thank you for all your feedback on my job market paper; it has helped me greatly in positioning my paper in the literature. In addition, I would like to thank Guia Bianchi, Tayler Bergstrom, and Professors Rakesh Sarin, Francesco Testa, Aimee Drolet and Timothy Malloy for their contributions on the first part of my thesis.

I am also very thankful to other faculty members at DOTM, especially Professors Felipe Caro, Fernanda Bravo, Scott Rodilitz and Francisco Castro. Felipe: thank you for sharing your feedback on my research and my job market talk; it was very helpful. Also thank you for being a great role model in teaching with the cases; I believe what I have learned from you will benefit me greatly in the future, while I will be teaching. Fernanda: thank you for the guidance during my first two years at UCLA when you were the Ph.D. liaison and for your feedback on my job talk. Scott: thank you for sharing your job market experience and giving me feedback on my job market talk; it helped me improve my presentation greatly. Francisco: thank you for sharing your feedback on my job market talk. I would also like to thank Craig Jessen for helping me overcome several challenges I faced throughout my Ph.D.; I knew I could always count on him for his guidance and support.

I was very lucky to share my Ph.D. journey with Jian Gao, Jingwei Zhang, Jingyuan Hu, Joon Kim, Pankaj Jindal, Saeed Ghodsi and Zach Siegel; thank you for all the great times we have shared. I owe a million thanks to İrem Akchen, Nur Kaynar and Yi-Chun Akchen. You became my family when I needed the most. Without your support and care, I could not have overcome all the challenges I faced in the past six years. I am looking forward to the great memories we will share in the future. I would like to thank Xinyi Guan, and Yilin Zhuo. My job market year would have been very lonely without you. Thank you for all the great memories we have shared. I would also like to thank Anna Saez for all her guidance and support both in my first and last year at Ph.D. Looking forward to become colleagues at IESE.

Outside of Anderson, I am very thankful to my dear friends, especially Cihan Karayazı, Hakan İşler, and Melis Saraçoğlu for their continuous support and all the great memories we have shared. You all made me feel like I am at home, when I was actually very far away from it. Thank you for all the late night talks and games. There was always a huge time difference between us but you were so accommodating that I always felt like you are right next door. I am looking forward to the many more great memories we will share. I would also like to thank Uğur Meci and Servet Bukce for becoming brothers that I never had and helping me throughout my life in Los Angeles whenever I needed.

I wouldn't be where I am today without the sacrifices my mother and grandmother, who unfortunately passed away during my Ph.D. journey, made for me throughout their lives. I am extremely lucky to have you as my family. I wish we could have spent many more years together. You have given me so much than what I could have ever wished for. I am very grateful for your endless trust and support in me. None of this would have been possible without you. I miss you so much in every moment of my life.

Last but not least, I want to thank my partner, Halid Çiçekli. Our paths crossed in a way we could never have predicted a few years ago. Meeting you was one of the best things that have ever happened to me in my life. Thank you for your love, companionship and continuous support; I don't know what I would have done without you. I look forward to building the next chapter of our lives together in Barcelona.

VITA

2016	B.S. (Industrial Engineering), Bilkent University
2018	M.S. (Industrial Engineering), Bilkent University
2018	Anderson Fellowship, UCLA Anderson School of Management

PUBLICATIONS

Yavuz M and Corbett, CJ (2024). Interactive Preference-based Optimization with Unknown Value Function: Application to Sustainable Sourcing in the Apparel Supply Chain. *Management Science* (under revision for resubmission)

Yavuz M, Bianchi G, Corbett, CJ, Bergstrom T, Drolet A, Malloy TF, Rajagopal D, Sarin RK, Testa F (2024). Interactive Preference-based Optimization with Unknown Value Function: Application to Sustainable Sourcing in the Apparel Supply Chain. *Journal of Industrial Ecology* (under revision for resubmission)

Argyris N, Karsu Ö, Yavuz M (2022) Fair Resource Allocation: Using Welfare-based Dominance Constraints . *European Journal of Operational Research* 297(2): 560-578.

CHAPTER 1

Introduction

Decision-makers have access to increasingly good information about the environmental consequences of their choices with the rise of life-cycle assessment (LCA) and related methodologies. These methods often implicitly assume that decision-makers make rational assessments when weighing environmental factors; however, what remains less recognized is the impact of behavioral biases and context effects, well-documented in the field of behavioral science, on their decisions. In addition, the current approaches in the sustainability literature provide little guidance on how to find an optimal solution when the decision-maker's preferences are unknown. Most work including environmental dimensions focuses on expressing environmental impacts in monetary values instead of directly trading off multiple environmental dimensions against one another. These trade-offs between sustainability-related dimensions are fundamentally different than traditional trade-offs involving one sustainability attribute and an economic attribute, as sustainability dimensions are often less tangible, more abstract, and more emotionally charged. Firms seeking to incorporate sustainability in their decisions face those difficult choices, with no well-defined guidelines on how to make trade-offs between different environmental and social impact categories. In this dissertation, our aim is to help decision makers with their choices in the context of sustainability by taking behavioral and optimization-based perspectives.

In the second chapter, we conduct a series of experiments and show that decision makers are equally vulnerable to context effects when facing choices involving trade-offs between environmental attributes as they are with conventional attributes. We delve into two specific context effects, namely the attraction and compromise effects, wherein the decision between two options may be influenced by a third, irrelevant option. Our work is framed in the context of making decisions based on the type of information that would be obtained from life-cycle assessment studies and is closely modeled on the way that LCA information is made available to designers in the apparel industry through the Higg Materials Sustainability Index, largely developed by the Sustainable Apparel Coalition.

Our research findings are based on a series of lab experiments conducted with diverse populations, including students and alumni from various environmental science programs at UCLA, alongside other student cohorts (via UCLA Behavioral Lab) and the general population (through MTurk). This diversity mirrors the array of backgrounds among decision-makers in real-world scenarios, aligning with our deliberate approach to encompass a wide range of perspectives. Our findings demonstrate that context effects, whether in environmental or non-environmental contexts, are frequently substantial and noteworthy. Our work highlights the importance of incorporating behavioral science into the broader environmental literature that involves trade-offs between environmental attributes, beyond its already established role in behavioral environmental economics.

In the third chapter, we propose an interactive optimization approach that can help decision makers with difficult trade-offs. We assume that we have an optimization problem that is linear, and the preferences of the decision maker are based on an unknown implicit linear value function. We propose a new interactive optimization approach that asks pairwise comparison questions to the decision maker to determine the solution that is most aligned with her preferences. Interactive optimization methods have been widely used in other settings, going back to the 1970s. We build on that body of work, but unlike existing methods, we do not base our algorithm on the adjacent efficient solutions or ask the decision maker for the acceptable amount and direction of trade-offs. Our algorithm elicits the preference information from the decision maker via pairwise comparison questions as needed and reduces the feasible region in each iteration. Our approach is also different from conjoint analysis, as in our case, the alternatives are not known in advance; instead, they are determined by an outer optimization problem during the algorithm, and it is impossible to present the decision maker with all comparisons due to the huge number of alternatives.

To illustrate the method, we use the context of sustainable sourcing in the apparel industry, as it accounts for substantial environmental impacts worldwide. We numerically test our approach using realistic data based on the Higg Material Sustainability Index (previously managed by the Sustainable Apparel Coalition) with three to four criteria: cost, global warming potential, water use and fossil fuel use. Interactive optimization has been surprisingly rarely used in the sustainability domain so far. Rowley et al. (2012) state that interactive optimization methods have not been observed in the environmental domain, and the recent literature review on operations research in sustainability assessment of products by Thies et al. (2019) also does not include any interactive approaches. In addition to conducting numerical experiments and benchmarking with a classical method by Zionts and Wallenius (1976), we reached out to several individuals working in the apparel industry to test our algorithm and received encouraging feedback.

During the interviews with several individuals working in the industry to test the method we developed in Chapter 3, we discussed how sustainable sourcing decisions are made in the industry. They said that such decisions are currently mostly made based on achieving some pre-set targets for the environmental attributes and/or using a total score for the environmental impact. One of the tools that are widely used in the fashion industry is the Higg Material Sustainability Index, which is a quantitative tool to help decision makers consider the environmental impacts of the raw materials used in apparel and footwear products based on extensive LCA data. They generate a total impact score to compare different raw materials.

Even though the total impact scores may be easier to use in the decision-making process, the total score will inevitably depend on some weighting aggregation of the environmental attributes. Making decisions based on such a total score might be problematic, as the weighting of the individual environmental attributes is very subjective and can change over time, even for the same decision-maker. To this end, it is important to have an interactive tool that captures the preferences of the decision maker in real time; however, it is also important to understand the reaction to this method and assess its practicality compared to other decision-making techniques currently in use.

In the fourth chapter, we develop an experimental framework to evaluate how participants react to the interactive preference-based optimization method, comparing it with a more conventional approach using an existing method called Direct Rating to elicit weights and then optimizing according to those weights. In the framework, participants will be exposed to both methods, and respond to a series of questions about which they prefer. The sequence in which they are exposed to the two methods will be randomized, and the effect of the ordering of the methods on their preferences can be tested. The framework also includes two treatments in which the sequence of the attributes involved is reversed (cost - global warming - water use, and water use - global warming - cost). This will allow us to measure whether participants weigh the first attribute more heaviliy, and whether any such sequencing effect varies between methods. We utilize oTree to create the experiments with the integration of Gurobi Web Licence Service and deploying it into an application on Heroku database.

This dissertation makes the following contributions: First, even though the literature broadly related to LCA tends to implicitly assume that decision makers are rational when making choices involving trade-offs between environmental attributes, we show that, to the contrary, decision-makers are equally at risk of falling prey to context effects (such as attraction or compromise effects) when facing choices involving trade-offs between environmental attributes as they are with conventional attributes. Our work highlights the importance of incorporating behavioral science into environmental disciplines such as life-cycle assessment. Second, we develop a novel interactive optimization approach applied in the domain of sustainable decision-making. The method we propose adds value especially when the decision problem is continuous and there are too many alternatives to present to the decision maker for her to make a rational decision. Since our method starts with an unknown underlying value function of the decision maker and learns about the decision maker's preferences in each iteration, it is a valuable tool for decision-making in sustainable operations. Third, we develop an experimental framework combining oTree with Gurobi optimization, a combination that has hardly been used in the operations management literature so far, but which could prove valuable for many other online experiments involving behavioral questions and optimization. Overall, this dissertation contributes to the literature on decision-making in sustainable operations by combining behavioral and optimization-based perspectives.

CHAPTER 2

Influence of Irrelevant Alternatives on Choices with Environmental Attributes

2.1 Introduction

Individuals in many settings have increasingly good access to information about various environmental impacts of their decisions, thanks to advances in life-cycle assessment (LCA) and other related fields, but this often exposes them to difficult choices where no alternative dominates the other. This is true for policymakers, executives and product designers within firms, consumers, and others. Making a data center more energy-efficient may increase its water use (Karimi et al. 2022). Increasing water efficiency in agriculture can reduce yield and increase land use (Pfister et al. 2011). Conventional cotton offers higher yield per hectare than organic, but at the expense of greater eutrophication, water consumption, and greenhouse gas emissions (Muthu 2020, p. 13). If adding odor-reducing nanoscale silver to textile reduces consumer laundering frequency, that would introduce trade-offs between various environmental and human health impacts (Hicks et al. 2015, Walser et al. 2011). The emerging field of alternatives assessment (Tickner et al. 2015) also typically involves making trade-offs between competing objectives. One approach to facilitate making choices in such settings is to aggregate the impacts, using some normalization and weighting scheme to help decision-makers rank the alternatives, as is common in LCA (Pizzolo et al. 2017), and also used in the Eco-Score front-of-pack environmental label that is being adopted by various retailers in Europe. Reap et al. (2008) discuss various problems associated with weighting, and aggregation is controversial and involves loss of transparency (Rajagopal et al. 2017), so others argue that it is better to treat the impact categories separately. Much of the literature on LCA and related methods tends to implicitly assume that decision-makers are rational when making trade-offs between environmental attributes. Linkov and Seager (2011) point out that, without a decision-analytic framework, LCA can leave decision-makers vulnerable to biases or cognitive limitations, and Pryshlakivsky and Searcy (2021) critically examine challenges associated with using LCA in a decision-making context. For example, consider the role of LCA in evaluating various biofuel alternatives, another domain involving trade-offs between emissions, water use, land use, and other attributes. McKone et al. (2011) discuss various challenges in this context, including whether the analyst should apply LCA to the full range of alternatives or only to a representative or an informative subset. We show that, as a result of cognitive biases, the analyst's choice of options to present can by itself already have a major influence on which alternative decision-makers select.

Why is it important to explicitly study trade-offs between environmental attributes? Environmental goals such as reducing climate change or reducing water use may have synergies but are also frequently in conflict with one another. A better understanding is therefore needed about how individuals make trade-offs between environmental attributes. Behavioral biases are usually examines between an economic and an environmental attribute, for instance willingness to pay to save habitat. When both attributes are environmental, choices require trade-offs between deeply held values or between attributes about which individuals have limited information, introducing conflicts and uncertainty in the mind of the decision-maker about their preferences. Compare that, for instance, with choosing an apartment and trading off between size and proximity to work, where individuals may have more determinate preferences based on their lifestyle choices. When preferences are uncertain, choices become even more difficult, and may be (even) more influenced by behavioral biases. We focus here on two specific biases, the attraction and compromise effects, both examples of context effects. An attraction effect occurs when adding an inferior alternative to a choice set increases the proportion of people who choose the dominating alternative from the original choice set; this would violate the regularity assumption of many rational choice models, which states that adding an alternative to a choice set should not increase the probability of choosing one of the original options. A compromise effect occurs if an alternative is chosen more frequently when it is a compromise rather than extreme option. Attraction and compromise effects have already been demonstrated in conventional contexts, often involving consumer choice; do these effects also occur when the attributes are environmental? Given the potential consequences of the type of environmental choices and policy decisions mentioned earlier, allowing context effects to cause a decision-maker to behave in a manner that is inconsistent with their underlying values can be costly.

We selected the apparel sector as the context for our experiments, as it is a major contributor to a variety of environmental (and social) impacts (Moazzem et al. 2022, Muthu 2020, Nielsen et al. 2022) and is relatively advanced in bringing LCA results related to material choices to decision-makers. Designers have a disproportionate impact on a garment's environmental impacts, through their selection of fabrics and materials. Policymakers have a similarly large impact when seeking to encourage or discourage use of certain materials through various forms of incentives or regulation. For our purposes, consider a designer working for an apparel firm, who faces a choice between two fabrics with different environmental attributes, where neither alternative dominates the other. Consider the two cases shown in Figure 2.1, which are based on the Higg Materials Sustainability Index, a tool that aims to make LCA results more accessible to practitioners in the apparel industry (more on which later).

In typical graphical representations of such choices in consumer behavior settings, alternatives that are higher or to the right are preferred, so throughout this chapter we reverse the axes in the figures; moving upwards or to the right means lower impacts and hence a preferred alternative. (Our experiments did not use such reverse-scaled figures.) As illustrated in the left panel in Figure 2.1, LCA studies suggest that acrylic fabric contributes more to

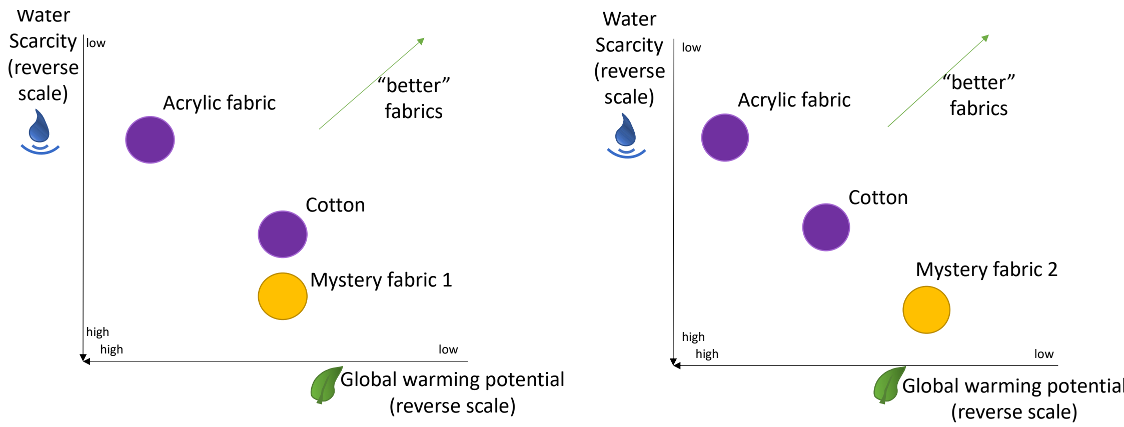


Figure 2.1: Example of attraction (on the left) and compromise (on the right) effects

climate change but less to water scarcity, while the reverse is true for cotton. Neither fabric is uniformly better than the other. However, one would presumably not want the choice between the two fabrics to be influenced by some third alternative fabric ("mystery fabric 1"), a decoy, which is dominated by cotton; this would be an example of the attraction effect. In the right panel, if a decoy third alternative ("mystery fabric 2") is added, some people might legitimately prefer that over acrylic and cotton; but people who prefer acrylic over cotton when those are the only two choices offered should not suddenly prefer cotton when a third option is provided, which would be an example of the compromise effect. In this study, we use a series of experiments to assess the frequency and magnitude of attraction and compromise effects when both attributes are environmental impacts, compared to a benchmark with non-environmental attributes.

As benchmark we selected a widely cited consumer behavior paper (Simonson 1989) that used a series of experiments to measure attraction and compromise effects in a variety of settings, such as trading off price and quality for a drink, or condition and distance from work for an apartment. We first replicated those experiments with the original non-environmental attributes. We then translated the choice sets of those experiments into equivalent choices with environmental attributes, in the context of designing garments. We found that the attraction and compromise effects occur approximately equally frequently with the environmental attributes as in our replication of the original experiments. Sometimes the choice frequency of an alternative changed by as much as 30 percentage points after adding a decoy, suggesting that context effects can have substantial and so far largely neglected effects on choice behavior in environmental settings.

The organization of this chapter is as follows. In section 2.2, we will provide a literature review that provides the background of our study. In section 2.3, we talk about the design of experiments and data collection. We also outline our statistical analyses. In section 2.4, we provide the results of our analyses both for the experiments with attraction and compromise effects. We provide descriptive statistics and discuss the effects of moderators as well. We conclude this chapter in section 2.5 with discussion and concluding remarks.

2.2 Background

There is a vast literature on biases that occur when trying to use incentives to encourage pro-environmental behavior, which typically involves trading off between one environmental attribute and an economic one (Kesternich et al. 2017, Osbaldiston and Schott 2012, Velez and Moros 2021), or, as Croson and Treich (2014, p. 347) write, "environmental economists typically study tradeoffs between the environment and other scarce resources (e.g., income)". Carlsson et al. (2021, p. 217) review the literature on the use of nudges as an environmental policy instrument, to "make it easier for the individual to "do the right thing". What if there is no "right thing"? Very few studies examine biases in trade-offs between multiple environmental attributes. Questions such as "how many tons of additional greenhouse gas emissions are an acceptable price for a given reduction in water footprint?" are critical for decision-makers of all kinds. Such questions permit no simple answer, but presumably we do not want the response to be heavily swayed by how the question is framed, such as whether dominated alternatives are eliminated before being presented to the decision-maker.

To date, there are very few studies on context effects involving decisions with conflicting environmental attributes. A few exceptions are Hämäläinen and Alaja (2008)'s experiments on the splitting bias in environmental choice, and the experiments in Mettier et al. (2006) and Mettier and Scholz (2008) on framing effects in LCA; Hämäläinen (2015) conceptually discusses biases in environmental modeling more generally. The closest to our work is Bateman et al. (2008), who report the existence of attraction (or asymmetric dominance) effects in the context of environmental management strategy options for a lake. They argue that the "demonstration of asymmetric dominance effects within choices for non-market environmental goods [... is ...] a potentially important result" (p. 125). Although behavioral scientists and economists will not be surprised by this finding, it does not seem to have generated much traction in the literature on LCA and related areas.

Despite the advances made in LCA, there is currently relatively modest understanding of how LCA is actually used in practice (Rex and Baumann 2008, Testa et al. 2016, Galindro et al. 2020). Baumann (2000) offers an early detailed description of LCA projects at two Swedish firms, and Hofstetter et al. (2002) and Hofstetter and Mettier (2003) conduct early surveys of users of an LCA-based tool for building products. The survey by Testa et al. (2016) of adopters and non-adopters of LCA among Italian firms pointed to data collection as a key hurdle to greater use of LCA; the Higg Materials Sustainability Index aims to address that obstacle. Galindro et al. (2020)'s survey of 55 practitioners reveals that they frequently use LCA information to choose between alternatives, despite citing "comparability issues" as the main disadvantage or weakness. Guérin-Schneider et al. (2018) report on their experiences working with the public wastewater sector in France to implement a simplified LCA calculator. Beemsterboer et al. (2020) suggest that simplification is "part of daily practice" for LCA practitioners and offer a systematic review of simplification strategies used.

During Spring 2019 we conducted exploratory interviews with about 10 individuals who used the Higg MSI to understand how they used the LCA-based information it provides; they typically did not use the "total impact score", consistent with the survey respondents in Hofstetter and Mettier (2003), 88% of whom reported not aggregating across impacts, most often arguing that such aggregation was not valid. Our interviewees also preferred to focus on the disaggregated impact categories; none of them systematically traded off between impacts, mostly focusing only on global warming potential. The value of tools such as LCA, and initiatives such as the Higg MSI that aim to make LCA results more accessible to practitioners, is to present them with credible and comparable information about the trade-offs associated with a choice they are facing; less is known about how they do or should make such trade-offs.

Much of the literature, including that on LCA, that focuses on providing the best possible estimates of environmental impacts implicitly assumes that behavioral biases do not occur. Our work examines whether these context effects do also occur in the case of environmental attributes. We focus on the attraction and compromise effects because they emerge naturally from the way LCA-based choices are presented, and because they are widely studied in non-environmental settings, examples include Huber et al. (1982), Simonson (1989), Frederick et al. (2014), Drolet et al. (2008), Drolet et al. (2020) and many more. We go beyond Bateman et al. (2008)'s earlier work on attraction effects in several ways: we benchmark our experiments involving environmental choices against a replication of a prior study involving conventional product attributes; we include various forms of compromise effects in addition to attraction effects; and we frame our experiments explicitly in the context of LCA.

2.3 Methods

In this section, we provide the general problem definition. The DM has an implicit value function over multiple attributes that she wishes to maximize. Following Dyer and Sarin (1979),

2.3.1 Design of Experiments

The goal of this work was to assess whether decision-makers are subject to context effects when making choices between alternatives when the attributes are environmental impacts rather than the more typical consumer-focused attributes related to price, quality, etc. To have a baseline against which to compare our results, we first replicated the 7 experiments reported in Simonson (1989)'s Study 1, which involved 372 students. Two of those experiments focused on attraction effects and five on compromise effects. We made a few minimal changes in the choice sets to make them more current and inclusive: we changed "beer" to "soda" (keeping the same attributes of price and taste), and we changed "calculator" to "cellphone" (changing one attribute from "number of functions" to "gigabytes of storage").

Subsequently, we constructed equivalent experiments mimicking choices made by apparel designers. We followed the structure of the Higg Material Sustainability Index (MSI), a tool that aims to make LCA results more accessible to practitioners in the apparel industry, but with hypothetical materials. We asked the participants to imagine they were apparel designers tasked with selecting among different blends of fabrics, involving trade-offs between the environmental attributes included in the Higg MSI (global warming potential, water scarcity, eutrophication, and abiotic resource depletion and use of fossil fuels). To enhance comparability between Simonson (1989)'s original experiments and our environmental replications, we kept the choices as similar as possible. We kept the relative proportions between the attribute scores the same as in Simonson's experiments, but we adjusted the scales to be centered around 100 and to ensure that higher scores represented higher impacts. For instance, the core set in Simonson's first experiment involved choosing between two sodas, one with a price of \$1.90 and a quality of 65, the other costing \$2.80 for a quality of 75. We translated that to choosing fabrics for a t-shirt, where the first had a water scarcity score of 76 and eutrophication score of 107, and the second scored 112 and 93 respectively. (The two decoys had water scarcity scores of 124 and 88, so the overall mean across the four alternatives is 100.) We included the definition of each of the environmental impacts at the beginning of the survey text presented to the participants. The information was presented in text and tables, not in the reverse-scaled graphical form shown in Figure 2.1. We included a few attention checks, and some demographic questions about factors such as age, gender, ethnicity, and environmental knowledge at the end of the experiment. We pre-registered the survey at aspredicted.org on July 6, 2020. The study was submitted to the university's Institutional Review Board with IRB#20-001029 and was certified as exempt from review on June 8, 2020.

We followed the structure of the Higg Material Sustainability Index (Higg MSI) as it is aimed at this population of apparel designers. Luo et al. (2021) include the Higg Index as one of the four main methods to evaluate environmental sustainability of textiles (alongside LCA, environmental footprint, and eco-efficiency), due to its "high potential in the commercial setting". See SAC (2020) for more on the underlying methodology, and the Appendix A.1 and Radhakrishnan (2015) for more background and history on the Higg MSI. As it becomes more established, the Higg Index is also attracting more scrutiny from the media and regulators; however, any limitations it may exhibit do not affect our findings, as our work does not depend on the actual data embedded in the tool.

2.3.2 Data Collection

All the data used in this study came from our experiments. Before conducting the experiments, we used a power analysis to estimate the desired sample size, based on the effect sizes found by Simonson (1989), and determined that a sample of 300 would likely be adequate. Our main target sample was participants in the Behavioral Lab (BLab) of a major US University. We planned to recruit additional responses through Amazon Mechanical Turk (MTurk) if we did not reach the desired sample size from the BLab; to compensate for lower average quality of responses from MTurk we planned to recruit 5 MTurk participants for every response that we fell short of 300 from BLab. In addition, we recruited students and alumni from the University's Sustainability Institute (SustI) as participants, in order to assess whether the results would differ among a more environmentally informed population. We removed responses that were incomplete or that failed the attention checks.

The initial BLab sample had 368 responses, of which 230 were retained for the analysis. This was 70 short of our target of 300, so we sought additional participants from MTurk. We obtained 288 MTurk responses, of which 260 were retained. We retained 148 out of 174 responses from SustI students (SustIS), and 15 out of 19 responses from SustI alumni (SustIA). The main analyses presented here are based on the full sample of 653 responses. Though the exact results do vary somewhat across samples, we did not detect any consistent patterns. The SustI respondents rated their environmental knowledge higher than other respondents, about 5.5 on a 7-point scale compared to 4.2 for the overall sample, but the context effects do not appear to be consistently smaller or larger among the SustI sample than among the BLab or MTurk samples. This suggests that environmental knowledge by itself does not mitigate these effects. We do obtain less statistical significance in the separate subsamples but that is to be expected given the smaller sample sizes relative to the full sample. We provide results for each sample separately in the Appendix A.

2.3.3 Statistical Analyses

We first inspected the data visually by looking at the choice proportions of each alternative under the various treatments. We found substantial evidence of patterns consistent with attraction and compromise effects. Subsequently we performed logistic regressions using R version 3.6.1. to determine if the attraction and compromise effects are statistically significant, after including various controls. The statistical analysis was independently replicated by a research assistant. Our regression equation (in shorthand) is as follows:

$$Y_i = \beta_0 + \beta_1 C_i [+\beta_2 (moderators)_i + \beta_3 C_i (moderators)_i] + \epsilon_i \tag{2.1}$$

The dependent variable $Y_i = +1$, if the target option is chosen and is 0 otherwise. The independent variable C_i indicates whether the treatment associated with response i involved a decoy or not. We used contrast coding (Judd et al. 2017) for C_i, i.e., $C_i = +1$ for the treatment with decoy and $C_i = -1$ for the core treatment. Using contrast coding rather than 0-1 dummy variables allows easier interpretation of the coefficient estimates. We included moderators such as age, gender, ethnicity, environmental knowledge, and environmental attitude, together with their interactions with the treatment variable. More details on the regression and results are provided in the Appendix A.

To isolate the attraction and compromise effects, we conditioned our analyses on the two options of interest. For the experiments on attraction effects, this means we eliminated the (very few) observations where the decoy option was selected. For the compromise effect experiments, several mechanisms can simultaneously influence how participants' choices shift between the core set and the treatment set. For instance, in the setting illustrated in the right panel in Figure 2.1, some participants may simply prefer to minimize GWP regardless of the other attribute, and hence will always choose the right-bottom option among those presented. Participants who choose the left-top option in the right panel in Figure 2.1 when presented with only two alternatives should not change their mind once a decoy (the right-bottom alternative) is added. To isolate the compromise effect in this case, we condition our results only on the non-decoy alternatives, excluding responses that selected the right-bottom (decoy) alternative. See below for an example in the attraction effect experiments, and the Appendix A.2 for an example in the compromise effect experiments.

2.4 Results

We observe statistically significant ($p < 0.05$) attraction and compromise effects, and with a similar range of magnitudes, in more than half of the experiments. They occur approximately equally frequently in our environmental translations of Simonson (1989)'s original experiments as in his original experiments.

We analyze the results visually and using logistic regression with a dependent variable indicating whether participants chose the target option and controlling for various other factors such as age, gender, environmental knowledge, and others. We summarize our findings below; more detailed descriptive statistics and regression results are provided in the Appendix A.

2.4.1 Descriptive Statistics

Table 2.1 and Table 2.2 summarize the key descriptive statistics for the overall sample with the sample size of 653; the Appendix A.3 shows the breakdown by subsample. Participants' age ranged from 16 to 78, skewing higher in the MTurk and SustIA subsamples. On a 7-point scale, the SustIS and SustIA groups reported higher environmental knowledge (5.33 and 5.61 respectively) than the BLab respondents (3.97), which in turn were higher than the MTurk respondents (3.68). Environmental attitudes were quite similar across all samples, varying from 5.98 (MTurk) to 6.50 (SustIA).

Continuous Moderator's Name	Min	Max	Mean	Std. Dev.
age (years)	16	78	29.36	12.51
environmental knowledge (scale 1-7; higher is more knowledgeable)	1	7	4.2	1.36
environmental attitude (scale 1-7; higher is more concerned)	2.9	7	6.1	0.78

Table 2.1: Descriptive statistics of continuous moderators for the overall sample

Gender		Ethnicity	
Female	384	American Indian or Alaska Native	5
Male	261	Asian	205
Other	4	Black or African American	28
Prefer not to say	4	Multiple	51
		Other	37
		White	327

Table 2.2: Descriptive statistics of categorical moderators for the overall sample

2.4.2 Attraction Effect

For the attraction effect, we conduct two experiments with two treatments each, for both the original and the environmental versions, leading to a total of 8 experiments. The attraction effect is measured by the difference in the choice proportion of the first option between a treatment with a decoy (irrelevant) option close to the target and a treatment without the decoy.

Consider the first two treatments, Soda I and its environmental counterpart T-shirt I. Figure 2.2 shows the choice proportions for both the core treatment and the treatment with decoy.

In Figure 2.2, the choice proportions in the core treatment without decoy are in illustrated with purple and those in the treatment with a decoy are illustrated in orange. The target option, which would benefit from adding a decoy if an attraction effect occurs, is indicated with an arrow. Alternatives to the right and higher in the figure are better.

In the core treatment 49% chose the target option B (higher price, higher quality). After adding a higher-priced but equal-quality decoy C, 77% of respondents now chose the target option B. To isolate the attraction and compromise effects, we use the relative choice proportions only of the alternatives that are present both in the core treatment and the

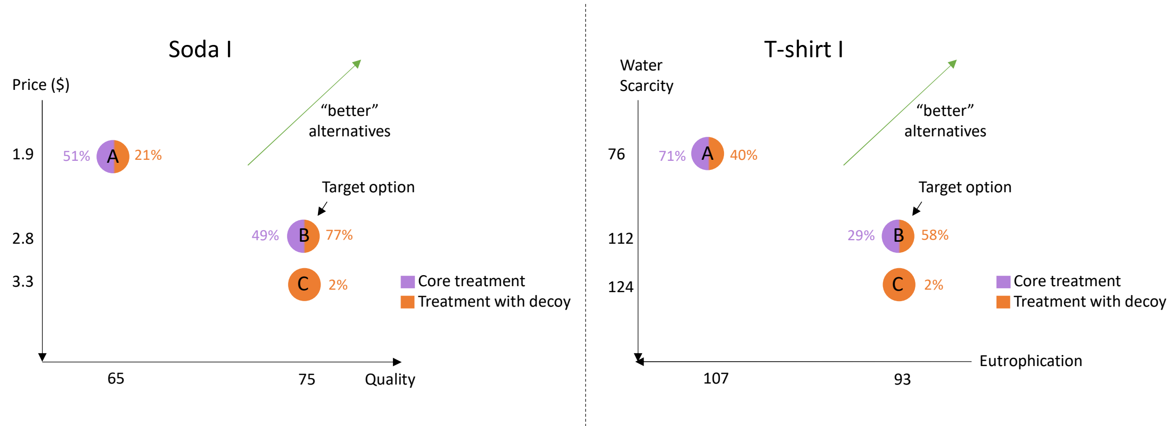


Figure 2.2: Absolute choice proportions for experiments Soda I and T-shirt I

treatment with decoy. In the Soda I case, that means ignoring the 2% that chose the decoy C. Conditional on choosing either A or B, 79% chose the target option B and 21% option A. The conditional choice proportion of the target option B increased from 49% to 79% when the decoy is introduced. This is statistically significant ($p < 0.05$) in the full logistic regression. In the equivalent environmental experiment, T-shirt I, 29% initially chose the target option (lower eutrophication, greater water use), which increased to 58% after adding the decoy C. The conditional choice proportion for B increased by 30 percentage points, again statistically significant ($p < 0.05$) in the logistic regression. In the next two experiments, Soda II and T-shirt II, the decoy was adjacent to option A, which becomes the target; no attraction effect was observed in the original experiment or its environmental version. (Figures are provided in the Appendix A.4.)

Simonson (1989)'s second experiment involved trading off ride quality against fuel economy for cars; we translated that to trading off between global warming potential (GWP) and water scarcity in designing jeans. A statistically significant ($p < 0.05$) attraction effect occurs in one of the two original treatments (Car I), and in the non-corresponding environmental setting (Jeans II). Figure 2.3 summarizes the conditional choice proportions of the target option in the four pairs of experiments for attraction effects. In Figure 2.3, purple bars show conditional choice proportions for the target option in the core treatment without decoy, orange bars represent the treatment with decoy. Choice proportions are conditional on choosing an alternative from the core set, excluding respondents who chose the decoy. When the orange bar is higher than the purple one, that reflects an increase in the conditional choice proportion of the target option when a decoy is added, i.e., an attraction effect occurred. Cases in which the effect is statistically significant ($p < 0.05$) in the full logistic regressions are marked with "*".

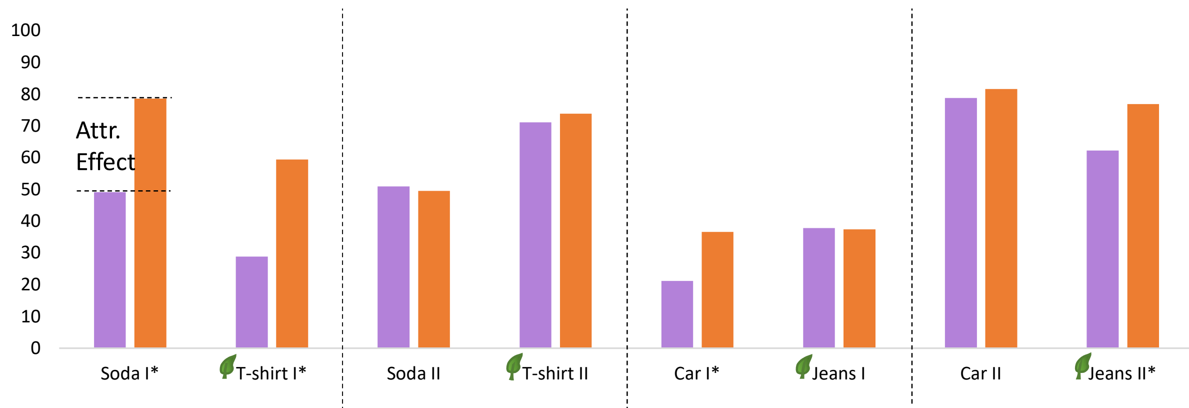


Figure 2.3: Conditional choice proportions for all attraction effect experiments

We conclude that significant attraction effects can occur as frequently and be equally substantial with environmental attributes as with conventional attributes, even if they do not occur in exactly the same cases.

2.4.3 Compromise Effect

For the compromise effect, we replicated 4 of the 5 experiments in Simonson (1989). (The fifth was coded incorrectly in the survey software; see Appendix A.7.) The target is the option that becomes a compromise after adding a decoy. Figure 2.4 shows the conditional choice proportions for these experiments. Similar to the Figure 2.3, purple bars show conditional choice proportions for the target option in the core treatment without decoy, orange bars represent the treatment with decoy. Choice proportions are conditional on choosing an alternative from the core set, excluding respondents who chose the decoy. When the orange bar is higher than the purple one, that reflects an increase in the conditional choice proportion of the target option when a decoy is added, i.e., a compromise effect occurred. Cases in which the effect is statistically significant ($p < 0.05$) in the full logistic regressions are also marked with "*".

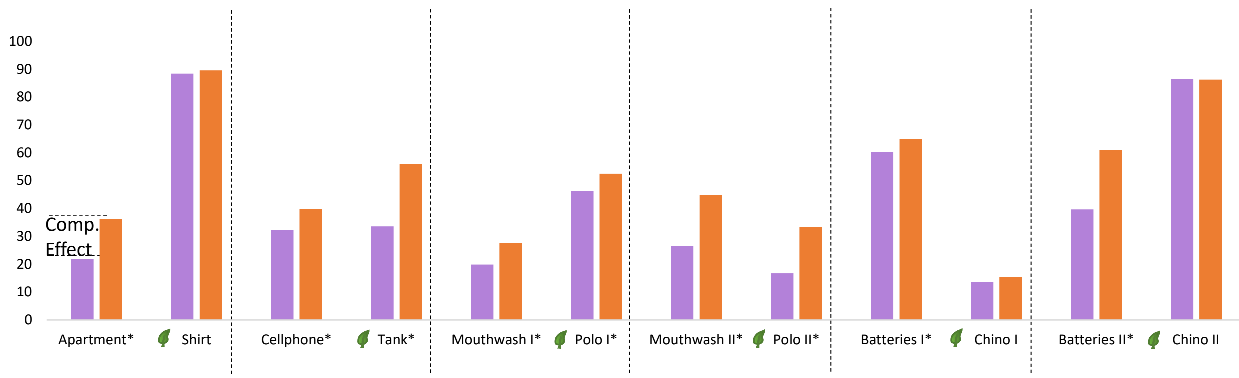


Figure 2.4: Conditional choice proportions for all compromise effect experiments

The overall pattern is comparable with that for attraction effects. Figure 2.5 shows the choice proportions for both the core treatment and the treatment with decoy for the Apartment and Shirt experiments.

In Figure 2.5, choice proportions in the core treatment without decoy are illustrated in purple, while those in the treatment with a decoy are illustrated in orange. Alternative D was shown, but it was not available to choose. The target option, which would benefit from adding a decoy if a compromise effect occurs, is indicated with an arrow. Alternatives to the right and higher in the figure are better.

For the Apartment experiment, participants either saw two alternatives, or the same two plus a decoy which was shown but not actually available to choose. In the core treatment,

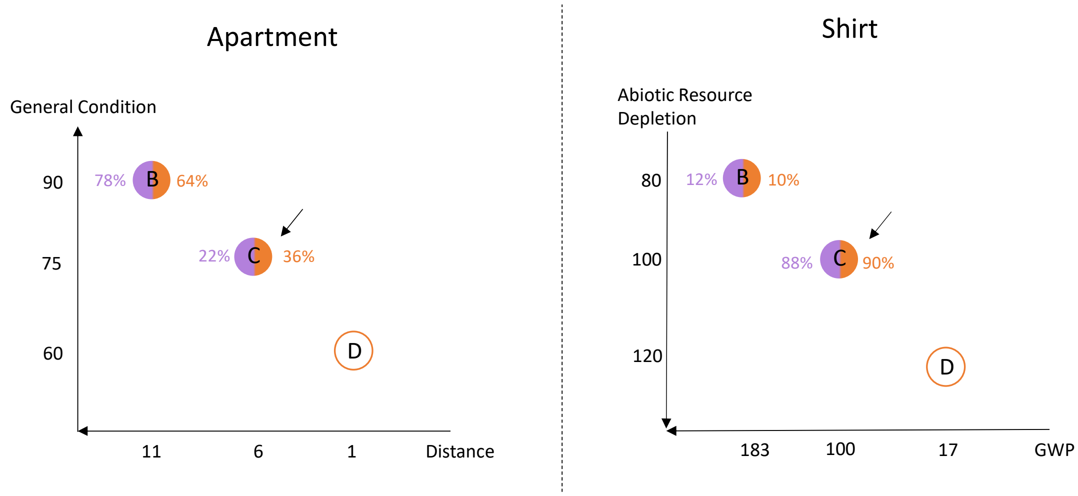


Figure 2.5: Absolute choice proportions for experiments Apartment and Shirt

22% chose the target option (closer to work but in worse condition). When adding the decoy (even closer to work but in worse condition), the choice proportion for the target increased to 36%, a statistically significant compromise effect ($p < 0.05$) in the logistic regression. In the environmental version of this experiment (choosing a fabric for a shirt), 88% already chose the target (lower GWP but higher abiotic resource depletion) in the core treatment; this increased further to 90% when the decoy was added. This increase is small and not statistically significant, as the choice proportion of the target was initially already very high, leaving minimal room for a compromise effect.

Figures for the remaining experiments are included in the Appendix A.4. In the Cellphone experiment, trading off between storage and reliability, the conditional choice probability of the target increased from 32% in the core treatment to 40% after adding a decoy, statistically significant ($p < 0.05$) in the regression. In the environmental version, designing a Tank top while trading off between water scarcity and GWP, the conditional choice proportion of the target increased from 33% to 56%, again statistically significant ($p < 0.05$). The two experiments involving Mouthwash (germ-killing effectiveness vs. fresh breath effectiveness) and the environmental counterparts with Polo shirts (abiotic resource depletion vs. water scarcity) show a similar pattern, and all yield statistically significant ($p < 0.05$) compromise effects. In the final experiments with Batteries (probability of corrosion vs. expected life) and with designing Chinos (GWP vs. eutrophication), the effects are smaller or non-significant. In the Chino experiments, 86% chose the low-GWP alternative in the core treatment, again leaving little room for a compromise effect, similar to the Shirt experiment.

The overall takeaway is similar to that for the attraction effect: significant compromise effects can occur equally frequently and with similar magnitudes in choices with environmental attributes as with conventional attributes. In all three cases where we did not observe a statistically significant compromise effect (Shirt, Chino I and Chino II), the vast majority of participants chose the low-GWP option when the other attribute was abiotic resource depletion or eutrophication. These results suggest that participants prioritize global warming over eutrophication and abiotic resource depletion regardless of the magnitude of the trade-offs involved (within the range of our experiments) but are willing to make trade-offs between global warming and water scarcity. Our experiments were not designed specifically to look for hierarchies between attributes, but this is an important direction to explore further as LCA results become more accessible to decision-makers.

2.4.4 Moderators

An overview of the full results of the logistics regressions for each experiment, including which moderators were significant, is shown in Appendix A.5. Our main conclusion is that no consistent effects appear to exist, even with a more relaxed significance threshold of $p < 0.10$. Older participants exhibited less strong context effects in 3 of the 20 cases (T-shirt I, Car I, and Cellphone) and stronger in 1 case (Polo I). Higher environmental knowledge was associated with stronger context effects in 2 cases. Stronger environmental attitudes is associated with stronger context effects in 3 cases, and weaker in 1.

A.13 in the Appendix A.5 reports the conditional choice proportions for each of the experiments for each subsample separately. While there are some differences between the subsamples, the overall patterns are quite consistent. Consider the first pair of experiments. In the overall sample, in the original experiment (Soda I), the choice proportion increased by 0.30 after adding a decoy, and the same increase was seen in the environmental replication (T-shirt I). The corresponding figures for the subsamples are 0.37 and 0.30 for BLab, 0.26 and 0.31 for MTurk, and 0.25 and 0.32 for the combined SustI samples. Perhaps most crucially, the pattern in the SustI samples is not systematically different from the other samples, suggesting that the more environmentally literate respondents are no less susceptible to context effects than the other participants.

2.5 Discussion and Conclusion

The literature broadly related to LCA tends to implicitly assume that decision-makers are rational when making choices involving trade-offs between environmental attributes. Here we show that, to the contrary, decision-makers are equally at risk of falling prey to context effects (such as attraction or compromise effects) when facing choices involving trade-offs between environmental attributes as they are with conventional attributes. Our work highlights the importance of incorporating behavioral science into environmental disciplines such as life-cycle assessment. Decision-makers (whether policymakers or product designers or others) need to be mindful of behavioral biases. Similarly, when sustainability analysts prepare information for decision-makers about the environmental performance of a range of alternatives, they need to be mindful that the way they present that information can severely influence the decision-makers' choice in ways that may not reflect their true preferences. For instance, should an analyst eliminate all dominated alternatives before presenting the final options to the decision-maker? One might argue that the analyst should always present all options, but our work shows that including dominated alternatives can have a substantial impact on the decision-maker's ultimate choice. We do not take a position here on whether the analyst should eliminate the dominated alternatives, but we draw attention to the fact that this decision is far from innocuous. If the decision-makers are consumers, our work suggests that they will also struggle to make meaningful choices when faced with environmental performance information along multiple dimensions. Can firms or policymakers use this to their advantage, and should they? A firm wishing to encourage customers to select its more climate-friendly product might nudge them to do so by adding a decoy product that is similar to but dominated by the climate-friendly option. Other well-documented biases will likely also influence choices. Environmental attributes such as greenhouse gas emissions are often difficult to interpret and decision-makers have limited intuition for the numbers or scales involved; common practices such as providing a comparison (such as describing a certain level of greenhouse gas emissions as "equivalent to driving an average car for 35,000km") might induce anchoring effects.

There is a substantial behavioral science literature on de-biasing, though no simple universal mechanisms appear to exist. For instance, the attraction effect appears to be weaker when information is presented visually rather than only numerically (Frederick et al. 2014). Adopting more formal structured decision-making approaches that highlight trade-offs and dominance clearly for decision-makers may also help in some cases (Beaudrie et al. 2021, Linkov and Seager 2011). Even the choice of which environmental attributes to consider in LCA is to some extent a value judgment and hence subject to bias (Grubert 2017).

Our work points to many possible further questions. We did not consider a hierarchy among the attributes, but in practice individuals may focus almost entirely on one attribute at the expense of another; for instance, they may choose a more climate-friendly alternative regardless of how much worse it is in terms of eutrophication. We see tentative signs of this in our experiments but our study was not explicitly designed to examine this. One could explore whether such hierarchies of environmental attributes exist, and whether they are invariant or context-specific. One could expand the set of attributes to include social factors, such as child labor or fair wages. It is also possible that individuals react much more strongly when harms decrease from "1" to "0" than from "2" to "1"; for instance, reducing greenhouse gas emissions from 100kg per unit to 0kg (possibly through the use of offsets) may influence choices much more than reducing them from 100kg to 1kg. There are many other biases and heuristics that could be explored in this context, such as anchoring, reference points, mental accounting, and more. Understanding the potential effects of such biases becomes all the more important when one recognizes that results of LCA typically are subject to uncertainty (Mendoza Beltran et al. 2018), which further exacerbates the extent to which cognitive limitations can lead to choices that do not reflect the decision-maker's true preferences.

CHAPTER 3

Interactive Preference-based Optimization with Unknown Value Function: Application to Sustainable Sourcing in the Apparel Supply Chain

3.1 Introduction

The fashion and textile industry produces more than 30 million tons of finished materials each year, generating more than 1,700 million tons of CO_2 emissions (Shen 2014, Global Fashion Agenda and Boston Consulting Group 2017). Cotton is the most profitable non-food crop in the world and is a primary raw material in the textile industry. Each kilogram of cotton — the amount required to produce one t-shirt and one pair of jeans — requires 20,000 liters of water (WWF 2020). Increasing awareness of sustainability among customers has contributed to rising eco-design practices in the fashion industry (Wang and Shen 2017). Several companies including H&M, Uniqlo, The North Face, Nike, Patagonia and New Balance incorporate sustainable practices (eco-design) into their supply chain designs (Shen 2014). Although much progress is still needed, the apparel industry is ahead of some others in terms of supply chain transparency (Dai and Tang 2022), a key element of sustainability.

Eco-design takes place in multiple stages of the supply chain such as material selection, supplier selection, production process choices and recycling. Supply chain managers who care about sustainability face difficult choices. For instance, when designing a product, material selection requires trade-offs among conflicting environmental impacts. Regular sheep wool results in higher global warming potential, but uses less water compared to recycled wool (SAC 2020). Cotton made in Africa results in less global warming potential compared to cotton made in China, India, Australia or the United States (SAC 2020). No clear guidelines exist on how to balance these sometimes conflicting environmental impacts. To make better decisions, material and supplier selection should be considered jointly; we call this the *sustainable sourcing problem.* We use this setting to motivate our proposed method, but our approach can be applied much more broadly.

We assume that the decision maker (DM) works at an apparel brand and has an implicit value function for cost and for several environmental attributes. She wishes to maximize this implicit value function, but is unable to formulate the function itself, as is usually the case in practice. How many tons of CO_2-equivalent greenhouse gas emissions are equivalent to one ton of water use? Exploratory interviews in 2019 with 10 users of the (then) Sustainable Apparel Coalition's Higg Material Sustainability Index (MSI) suggest that they avoid such trade-offs and largely focus on a single environmental attribute. Further preliminary conversations, conducted in 2023 as part of this current work, including with practitioners at the kind of firms mentioned above, confirm that no formal methods are currently in use to quantitatively incorporate such trade-offs. Erhun et al. (2021) also comment that most companies still treat sustainability as a constraint – hence avoiding the need to be explicit about trade-offs between attributes – and argue that it should become one of the drivers of supply chain performance. Accommodating a DM's unknown implicit value function over multiple environmental attributes constitutes our main challenge.

We assume that the DM can correctly give preference orderings associated with different alternatives based on her implicit value function, but is unable to articulate that value function explicitly. If the DM has extensive access to a consultant with expertise in decision analysis and optimization, various customized approaches might be feasible, possibly eliciting initial weights and refining them iteratively while optimizing. Beaudrie et al. (2021) and the references cited there point to the importance of facilitation when facing multi-criteria decision problems. However, in many practical settings, a decision maker has to make such decisions alone, only aided by software. This is the context we study. The current approaches in the sustainability literature provide little guidance on how to find an optimal solution when preferences are unknown. Most work including environmental dimensions focuses on expressing environmental impacts in monetary values rather than directly trading off multiple environmental dimensions against one another. The current adaptations of approaches in the multi-criteria decision making literature to the sustainability domain do not necessarily return the optimal solution for a given DM, because either they focus on satisficing rather than optimizing by introducing a stopping criterion allowing the DM to terminate when an acceptable solution is found, or they elicit weights a priori and then assume the implicit value function of the DM is known while optimizing, which can lead to highly suboptimal solutions if the weights were not elicited correctly. Some approaches find the set of Pareto efficient solutions and present those to the DM, but for a complex optimization problem with many feasible solutions, even with three criteria, this will be impossible for a DM to digest.

To motivate and illustrate our method, we consider sustainable sourcing in the fashion industry, where the environmental preferences of the DM are unknown. We propose an interactive optimization approach that asks pairwise comparison questions to the DM to determine the sourcing mix that is most aligned with her preferences. We aim to do so by requiring minimum cognitive effort from the DM by asking as few questions as possible that are also relatively easy to answer. Using a series of pairwise choices to find a DM's optimal solution without knowing the value function dates back to the work of Zionts and Wallenius (1976) and several others since then; the way we generate the alternatives to be presented is different and has promise. We compare our approach to Zionts and Wallenius (1976) and others in the literature section. Although we focus on the sustainable sourcing problem, the interactive optimization approach can be applied to many other settings with unknown implicit value functions with multiple objectives. Many of the examples discussed in Jónasson et al. (2022) and Sunar and Swaminathan (2022) also involve trading off multiple social and environmental attributes.

Our main contributions for this chapter can be summarized as follows. To the best of our knowledge, this is one of the first interactive optimization approaches applied in the domain of sustainable decision-making. Rowley et al. (2012) state that interactive optimization methods have not been observed in the environmental domain, and the recent literature review on operations research in sustainability assessment of products by Thies et al. (2019) also does not include any interactive approaches. Interactive optimization methods have been widely used in other settings, going back to the 1970s. We build on that body of work and use several elements of existing methods. The specific combination of models we use to generate alternatives for the DM to rank differs from prior work; among others, unlike some existing methods, we do not base our algorithm on the adjacent efficient solutions or ask the DM for the acceptable amount and direction of trade-offs. Our algorithm elicits the preference information from the DM via pairwise comparison questions as needed and reduces the feasible region in each iteration. Although reminiscent of conjoint analysis, our approach is different as the alternatives are not known in advance; instead, they are determined by an outer optimization problem during the algorithm, and it is impossible to present the DM with all comparisons due to the huge number of alternatives. If the underlying implicit value function is a convex combination of the objectives and the feasible region is a compact polyhedron, then our algorithm converges to the optimal solution. To test our algorithm, we constructed numerical experiments based on data adapted from the Higg MSI (formerly owned by the Sustainable Apparel Coalition or SAC). The algorithm converges to the optimal solution in a reasonable number of questions, with minimal computation time required at each iteration; in our experiments with three criteria, within on average around 15 questions. This is less than the classic Zionts-Wallenius (Zionts and Wallenius 1976) method, and most of the choices the DM has to make in our method are also easier as the alternatives are more distinct.

The interactive preference-based optimization method we propose adds value when the decision problem is continuous and there are too many alternatives to present to the DM for her to make a rational decision. If the underlying value function of the decision maker is known in advance with certainty, then our method is not needed, as one can just use those known weights and optimize the decision problem accordingly. However, that is rarely the case in real life and it is often difficult to elicit the weights of the DM, especially when the attributes are less tangible, more abstract, and more emotionally charged, as with problems involving environmental or social aspects. Since our method starts with an unknown underlying value function of the DM and learns about the DM's preferences in each iteration, it can be a valuable tool for such decision contexts in sustainable operations. In addition, our method can work directly with the DM without needing an analyst to guide the process, as it is fully automated.

We conducted preliminary interviews with 8 individuals including 5 from the apparel industry, working in sourcing and/or sustainability departments of their firms, and asked them to solve a hypothetical sourcing problem using our proposed method. They told us that no formal methods are currently used in practice for such problems involving trade-offs between multiple attributes. They suggested that in addition to finding the optimal solution, our method also has the potential to help firms gain a better understanding of their own values and decisions.

This chapter is organized as follows. In Section 3.2, we review related literature and outline our contributions. In Section 3.3, we provide the general formulation to our problem, and in Section 3.4 we introduce the interactive optimization approach. In Section 3.5, we provide the results of our numerical experiments on the sustainable sourcing problem. In Section 3.6, we summarize our conversations with practitioners about the method. We discuss our overall findings, conclusions and limitations in Section 3.7.

3.2 Literature Review

In this section, we discuss how our work relates to and differs from existing research. We first provide some background on sustainable sourcing in the apparel industry, including Life Cycle Assessment (LCA), to explain why the time is ripe for methods such as that proposed here to be used in sustainable sourcing. We describe how our method relates to conjoint analysis and then review literature on sustainable sourcing with multiple criteria. We conclude with a review of interactive algorithms in multi-objective optimization, on which our work builds.

3.2.1 Sustainable Sourcing in the Apparel Industry

Sustainability is a growing theme in the apparel industry for both scholars and practitioners (Shen 2014, Morana and Seuring 2011). Some of the environmental impacts associated with excess production can perhaps be mitigated by mass customization (Alptekinoğlu and Örsdemir 2022), but the other main impacts are those associated with materials. Most work in this field focuses on case studies of the leading textile firms. For instance, Wang and Shen (2017) analyse product-line data of Patagonia and Shen (2014) studies the supplier selection mechanism at H&M. Bevilacqua et al. (2014) report that the environmental impacts associated with the cotton yarn supply chain vary widely between different suppliers in different countries. Although optimization-based approaches are uncommon in this domain, Tseng and Hung (2014) propose an optimization model to minimize the sum of the operational costs and social costs caused by carbon dioxide emissions. They convert the emissions into monetary values and minimize the sum of all costs; our approach does not convert environmental impacts into monetary values but focuses directly on the trade-offs between them, and does not assume the relative costs or weights are known.

LCA is a quantitative method that evaluates a product system's inputs, outputs and the potential environmental impacts throughout its life cycle (Guinée and Heijungs 2017).

LCA studies in the textile industry mainly focus on determining the environmental impact of the fibers. For instance, Muthu et al. (2012) developed an environmental impact index and a related ecological sustainability index for a wide range of textile fibers. Similarly, van der Velden et al. (2014) worked on the environmental burden of textiles made of some base materials to determine which material and life-cycle stage has the greatest impact. Sandin et al. (2013) conducted an LCA case study of a wood-based textile fiber to assess water and land use impacts of bio-based textile fibers. LCA studies sometimes aggregate all environmental impacts into a single score using various weighting schemes, but these are based on environmental burdens rather than on the DM's preferences. Testa et al. (2016)'s survey of Italian firms finds that data collection is a key hurdle to greater adoption of LCA; the Higg Materials Sustainability Index (MSI) aims to address that obstacle. The Higg MSI is increasingly widely used in the apparel industry. It uses information from LCA studies and industry sources to provide users with quantitative information to allow them to compare the environmental impact of a wide range of materials used in apparel in four impact categories: climate change, eutrophication, abiotic resource depletion / fossil fuels, and water resources depletion / scarcity. See SAC (2020) for more on their methodology, and Radhakrishnan (2015) for more background and history. Although initiatives such as the Higg MSI provide practitioners who are not experts in sustainability with greater access to sustainability-related information, the LCA literature provides no other guidance on how to optimize a system or make trade-offs with these data; our interactive decision aid algorithm aims to contribute to this interpretation stage of LCA.

3.2.2 Sustainable Sourcing with Multiple Criteria

Existing work on sustainable sourcing with multiple criteria focuses mainly on problems where the alternatives are known in advance, i.e., choice problems. For instance, Memari et al. (2019) propose a fuzzy approach to determine the rank of sustainable suppliers. They ask the DM to rate the importance of each criterion and then assign ratings to each supplier.

Similar works include Büyüközkan and Çifçi (2011), Govindan et al. (2013), Öztürk and Özçelik (2014) and Alavi et al. (2021). In optimization-based approaches, i.e., multiple objective optimization, the alternatives are not known in advance; they are determined by the optimization problems. Manzardo et al. (2014) minimize the weighted sum of cost and water footprint of chemical pulp. They determine the weights first using a similar approach as in the choice problems and then optimize assuming those weights are correct. Similarly, Mierlo et al. (2017) study the environmental impacts of meat replacers, using two different approaches: minimizing the sum of the normalized impacts, and minimizing the maximum impact. As Thies et al. (2019) also point out in their extensive literature review on multiple-criteria decision making applications for sustainability assessments of products, none of the studies deploying multi-objective optimization within this context are interactive.

3.2.3 Conjoint Analysis

Our method is reminiscent of conjoint analysis, widely used in marketing and product design. Toubia et al. (2004) develop a polyhedral choice-based conjoint analysis method used to design questions. Their method adapts the next question to ask based on the previous answers. They perform Monte Carlo simulations to evaluate their proposed polyhedral method. Toubia et al. (2003) compares the polyhedral methods to the existing ones. Bertsimas and O'Hair (2013) consider a similar algorithm to learn the preferences of the DM and use robust optimization. They also provide a method to handle inconsistent preference information. Our algorithm also introduces hyperplanes to cut the feasible region as it learns the preferences. The main difference is that conjoint analysis works well when the alternatives are known in advance and/or there is a limited number of possible comparison questions. Our algorithm is designed for the case where there are many alternatives so that it is not possible to ask every comparison question, and it does not require knowing any of those alternatives in advance as they will be determined after solving corresponding optimization problems.

3.2.4 Optimization-based Interactive Algorithms

Multiple objective optimization techniques differ according to whether the timing of the articulation of preferences is a priori, interactive or a posteriori. The examples above were a priori approaches, while our work falls in the interactive category; to the best of our knowledge, optimization-based interactive algorithms have not yet been applied to sustainable operations problems such as sustainable sourcing. As environmental and social information becomes more available, due to advances in LCA and other methods, we believe the time is approaching for such methods to be more widely used in sustainable operations. Our method seeks to offer an improvement over existing methods for problems that are linear or the many others which can be approximated by linear models (Korhonen et al. 2012); some of the other methods reviewed below are more suited for other types of multi-objective optimization problem with unknown value function.

The literature on optimization-based interactive algorithms in multiple criteria decision making in other domains is well-established in MS/OR and dates back to the 1970s. Benayoun et al. (1971) introduce the STEP Method, which minimizes the maximum distance to an ideal infeasible solution. At each iteration, the DM is asked to specify the criterion to worsen to improve the others by an acceptable amount. Geoffrion et al. (1972) introduce an interactive method based on the Frank-Wolfe algorithm. They assume the utility function is concave and differentiable in the objectives over a convex and compact feasible region. They ask the DM to provide an estimate of her marginal rates of substitution (MRS) among the criteria, as well as the direction to proceed. Similarly, Oppenheimer (1978) assumes the preference function, which is either the sum-of-exponentials or sum-of-powers, is continuously differentiable, concave, non-satiated and has decreasing MRS and the feasible set is convex and compact. His method also assesses trade-offs and terminates when the DM cannot find any feasible improvement.

An experimental study by Wallenius (1975) evaluates the performance of the STEP

Method by Benayoun et al. (1971), the Geoffrion-Dyer-Feinberg (GDF) Method by Geoffrion et al. (1972) and a trial and error procedure introduced by Dyer (1973). In the experiment, subjects, who either belong to the manager or the student group, were asked to evaluate the methods according to their confidence in the solution, ease of use, ease of understanding and usefulness of information and were asked to provide their preference rankings of the methods. The findings indicate that the GDF method scores more poorly because the subjects had difficulty assessing the marginal rates of substitution. The unstructured trial and error method was the most favored one, mainly due to ease of use. As the above methods were hard to use in practice, Zionts and Wallenius (1976) introduce an interactive algorithm — the Zionts-Wallenius (ZW) method — to perform better in practice. They assume an additive linear implicit utility function. Their method is similar to the simplex algorithm but the decision maker decides which variable should enter the basis by specifying preference information.

Others have built on the ZW method over the years. Zionts and Wallenius (1980) discuss the technical properties of the models used and Zionts (1981) and Korhonen et al. (1984) modify the ZW method for discrete alternatives. Zionts and Wallenius (1983) modify the ZW method for a pseudo-concave utility function and propose checking all the adjacent solutions, which converges to a local optimum; further analysis is required to find the optimal solution. Köksalan and Sagala (1995) develop an interactive algorithm for discrete alternative multiple criteria decision making problems assuming a monotonic utility function. One of our main contributions is the incorporation of a "challenger model" which checks whether a strictly better solution exists at each iteration.

In all the above works, possible inconsistencies in the DM's responses are either ignored or treated by eliminating the old responses assuming DM learns more about the problem as the approach proceeds. Phelps and Köksalan (2003) propose an interactive evolutionary meta-heuristic for multi-objective combinatorial optimization problems. Pairwise comparison questions are asked to the DM and when an inconsistency is detected, they eliminate the responses in the optimization framework but they preserve the ordering while searching for a challenger. In this chapter we assume the DM's responses are consistent with this value function, but future work can explore extensions allowing for inconsistencies. Shin and Ravindran (1991) provide an extensive literature review on the early interactive multiple objective optimization methods. Dyer et al. (1992) provide an extensive review on multi-criteria decision making in general, which is later updated by Wallenius et al. (2008). Köksalan and Wallenius (2012) also provide a summary of the foundational methods and concepts of MCDM research.

Several more recent papers propose further variations, each with their own pros and cons. Deb et al. (2010) develop an evolutionary multi-objective optimization algorithm, where they assume a strictly monotone value function and test it on unconstrained problems. Mackin et al. (2011) propose an interactive weight space reduction procedure. They assume convex feasible region and concave, continuously differentiable objective functions, and they use its linear approximation for simplicity and in each DM call, they ask for preference information over multiple solutions. Sinha et al. (2014) introduce an evolutionary algorithm combining polyhedral cones and restrict the number of DM calls, where in each call they ask the DM to choose the best alternative among a number of undominated solutions. Korhonen et al. (2016) revisited their earlier idea in Korhonen et al. (1984) and introduce dual cones making the previous results computationally more tractable. Lokman et al. (2016) develop an interactive algorithm for multi-objective integer programming problems, where they utilize convex cones and assume a quasi-concave value function. They add additional binary constraints and variables in each iteration, which requires substantial computational effort. Later in Lokman et al. (2018) they propose a heuristic based on approximate cones to reduce the computational effort.

Toffano et al. (2022) propose a similar method for a combinatorial mixed-integer linear problem, but take a more AI-based perspective generating alternatives to present to the DM that they call "active learning". The visual description of this approach in their Figure 1 applies to our approach too, as well as to many of the other earlier works mentioned here. In their Figure 1a they refer to a sequential approach consisting of first generating weights and then solving the problem as the “state of the art”, which appears at odds with the extensive literature in this field (of which we only cite a selection here). It may create the impression that the concept of an iterative algorithm, such as their “proposed approach”, based on iterating between learning about a DM’s preferences by posing pairwise choices to generate weights and solving the problem with updated sets of weights, is more recent than it actually is. (Most of the work we review here on multi-objective optimization with implicit value function is indeed not cited in Toffano et al. (2022).) The novelty of the approach in Toffano et al. (2022) lies in the specific method used to generate alternatives, not in the overall structure of their framework.) They comment that their method might need to engage in too many interactions with the user for only marginal incremental improvement, so they introduce a stopping criterion to ensure termination.

Many of the above methods seek to generalize their applicability to non-linear value functions, convex feasible regions or integer programming problems. However, these all come with some sacrifices in practical performance, such as increased computational time and complexity, requiring too many DM calls, asking cognitively more demanding preference questions, or loosening the exactness of the methods by allowing termination at sub-optimal solutions. The interviews we conducted with industry practitioners confirmed the need for fast and easy to use approaches. They also highlighted the importance of trusting the method and its outcome, so having a tool that is intuitive and converges to the optimal solution would be beneficial in building that trust. Our algorithm aims to serve that purpose. Future work can explore combinations of elements of our approach with elements of these other methods to deal with a broader class of problems.

Our work is closely related to the ZW method, as our algorithm also reduces the feasible region by asking pairwise comparison questions. Our method generates pairwise comparisons that are not based on adjacent solutions, which means that the comparisons can be more distinct and potentially cognitively less demanding and also that they cover larger parts of the solution space, which potentially leads to quicker convergence and less number of questions. We use the ZW method as a benchmark and discuss the relationship between our approach and the ZW method in more detail in Section 3.5.4.

3.3 Problem Definition and Model Formulation

In this section, we provide the general problem definition. The DM has an implicit value function over multiple attributes that she wishes to maximize. Following Dyer and Sarin (1979), we refer to the preference representation function as value function instead of utility function, since we do not consider lotteries. She can correctly provide preference orderings using this implicit value function, but she is unable to articulate the function itself.

Assume there are K criteria and $k \in \mathcal{K} = \{1, ..., K\}$ is the set of criteria. (We present the model in general terms, but our method is most suited to 3 or maybe 4 criteria.) A decision variable vector $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ is associated with objective values $f_k(\mathbf{x})$ for $k \in \mathcal{K}$, which can represent the quantity of each material sourced from different suppliers and the functions $f_k(\mathbf{x})$ can be the resulting cost, global warming potential (GWP), and water usage, respectively. The set $\mathcal{X}$ denotes the feasible region, which we assume is a compact polyhedron. The DM optimizes $\mathbf{x}$ over $\mathcal{X}$ with respect to an implicit value function $U = U(\mathbf{f(x)})$, where $\mathbf{f(x)} = (f_1(\mathbf{x}), \ldots, f_K(\mathbf{x}))$. Thus, the general structure of the problem can be written as follows:

$$\begin{aligned} \max_{\mathbf{x}} \quad & U(\mathbf{f(x)}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X} \end{aligned} \tag{3.1}$$

Let $\mathbf{y}$=$\mathbf{f(x)}$. We assume each objective $y_k = f_k(\mathbf{x})$ is linear in $\mathbf{x}$, and that the implicit value function $U(\mathbf{f(x)}) = U(\mathbf{y})$ is a convex combination of these objectives, i.e., $U(\mathbf{y}) = \mathbf{w}^T\mathbf{y}$, $\mathbf{w} \geq 0$, $\sum_{k=1}^{K} w_k = 1$, where the vector $\mathbf{w}$ denotes the unknown weights associated with the objectives. Korhonen et al. (2012) conducted an experimental study on bi-criteria problems and observed that a linear value function seems to predict choices quite well.

Weak preference is denoted by $\preceq$, while $\prec$ and $\sim$ denote strict preference and indifference, respectively. We also assume that the decision maker is rational, i.e., preferences are complete (for alternatives $\mathbf{y^1}$ and $\mathbf{y^2}$, either $\mathbf{y^1} \preceq \mathbf{y^2}$, $\mathbf{y^2} \preceq \mathbf{y^1}$, or both must hold) and transitive (for alternatives $\mathbf{y^1}$, $\mathbf{y^2}$ and $\mathbf{y^3}$, if $\mathbf{y^1} \preceq \mathbf{y^2}$ and $\mathbf{y^2} \preceq \mathbf{y^1}$, then $\mathbf{y^1} \preceq \mathbf{y^3}$). For two vectors of alternatives $\mathbf{y^1}$ and $\mathbf{y^2}$, if $\mathbf{y^1} \preceq \mathbf{y^2}$, then by definition, $U(\mathbf{y^1}) \leq U(\mathbf{y^2})$, i.e., $\mathbf{w}^T\mathbf{y^1} \leq \mathbf{w}^T\mathbf{y^2}$. (Without loss of generality we assume maximization.)

Denote $\mathcal{P}^{\preceq}$ as the set of all ordered pairs of feasible solutions, i.e., $\mathcal{P}^{\preceq} = \{(\mathbf{y^n},\mathbf{y^p}) \mid \mathbf{y^n} \preceq \mathbf{y^p}\}$, which fully characterizes the DM's preferences, where the superscripts n and p refer to non-preferred and preferred. We say that a weight vector $\mathbf{w}$ is $l$ogically consistent with $\mathcal{P}^{\preceq}$ if and only if $\mathbf{w}^T\mathbf{y^n} \leq \mathbf{w}^T\mathbf{y^p}$ for all pairs $(\mathbf{y^n},\mathbf{y^p}) \in \mathcal{P}^{\preceq}$. Given $\mathcal{P}^{\preceq}$, we can jointly infer a feasible value of $\mathbf{w}$ and optimize $\mathbf{x}$ with the following bi-linear formulation:

$$\max_{\mathbf{w},\mathbf{x},\mathbf{y}} \quad \mathbf{w}^T\mathbf{y}$$

$$\text{s.t.} \quad \mathbf{x} \in \mathcal{X} \tag{3.2}$$

$$\mathbf{y}=\mathbf{f}(\mathbf{x}) \tag{3.3}$$

$$\mathbf{w}^T\mathbf{y}^n \leq \mathbf{w}^T\mathbf{y}^p \quad \forall\, (\mathbf{y^n}, \mathbf{y^p}) \in \mathcal{P}^{\preceq} \tag{3.4}$$

$$\sum_{k=1}^{K} w_k = 1 \tag{3.5}$$

$$\mathbf{w} \geq 0 \tag{3.6}$$

Constraint (3.2) forces $\mathbf{x}$ to be chosen from the feasible set $\mathcal{X}$; (3.3) defines the values of the objectives; (3.4) ensures that weight vector $\mathbf{w}$ is consistent with the DM's preferences; and (3.5) and (3.6) ensure that implicit value function's weight vector $\mathbf{w}$ satisfy the conditions of a convex combination. In practice, the preference set $\mathcal{P}^{\preceq}$ is unknown a priori and gathering pairwise comparisons for all feasible alternatives is highly inefficient; the number of possible comparisons is finite when $\mathbf{f}(.)$ and $U(.)$ are linear but can quickly become prohibitive.

Hence, we need a solution approach that can efficiently solve this problem without knowing the complete set of preference orders.

3.4 Interactive Preference-based Optimization

In this section, we propose an interactive preference-based optimization method that solves the bi-linear formulation efficiently. It consists of two main procedures. The first finds an optimal alternative $\mathbf{y}= \mathbf{f(x)}$ given a weight $\mathbf{w}$ consistent with the DM's preferences as revealed so far. The second procedure interacts with the DM and gets her preference information by asking a pairwise comparison question. Then it generates a new weight vector consistent with the updated preferences. The two procedures are applied iteratively until we obtain an optimal solution $\mathbf{y^*}=\mathbf{f(x^*)}$ or reach the termination criterion. The method is guaranteed to converge to an optimal solution in a finite number of steps, but one could choose to terminate earlier if the incremental improvements are too small. We introduce the optimization models used in the algorithm in Section 3.4.1, describe the overall algorithm and provide the pseudo-code in Section 3.4.2, and provide basic theoretical properties in Section 3.4.3.

3.4.1 Optimization models used

Let $\mathcal{P}_i^{\preceq}$ denote the set of ordered pairs $(\mathbf{y^n}, \mathbf{y^p})$ such that $\mathbf{y^n} \preceq \mathbf{y^p}$ according to the preference information gathered from the DM until iteration i, where i represents the number of questions asked to the DM so far. We first decompose the bi-linear formulation in Section 3 into two linear programming models: the main model (MM) and the feasibility model (FM).

$$\begin{aligned} \text{MM}(\mathbf{w}) \ \max_{\mathbf{x}} \quad & \sum_{k=1}^{K} w_k f_k(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X} \end{aligned} \tag{3.7}$$

$$
\begin{aligned}
\mathrm{FM}(\mathcal{P}_i^{\preceq}) \quad \max_{\mathbf{w}} \quad & 0 \\
\text{s.t.} \quad & \sum_{k=1}^{K} w_k y_k^n \leq \sum_{k=1}^{K} w_k y_k^p \quad \forall\, (\mathbf{y^n}, \mathbf{y^p}) \in \mathcal{P}_i^{\preceq} && (3.8) \\
& \sum_{k=1}^{K} w_k = 1 && (3.9) \\
& w \geq 0 && (3.10)
\end{aligned}
$$

The main idea is as follows: in the ith iteration, we first ask the DM a pairwise comparison question. We then include the new preference information in set $\mathcal{P}_i^{\preceq}$ and find new feasible weights $\underline{\mathbf{w}}$ that are consistent with all the preference information gathered so far by solving the feasibility model FM($\mathcal{P}_i^{\preceq}$). This update reduces the feasible weight space so it can lead to a reduction in the number of alternatives that can be optimal for the DM. If we are able to find weights that are consistent with the DM's updated preferences, then we can find a *challenger* to the current preferred alternative by using the new weights $\underline{\mathbf{w}}$. We define the *challenger* $\underline{\mathbf{x}}$ as the optimal solution to MM($\underline{\mathbf{w}}$), in other words, it "challenges" the most preferred solution of the DM so far. This structure is not new, but our contribution lies in the addition of a supplemental "challenger model" to collectively ensure a challenger is found, if one exists, even if MM($\underline{\mathbf{w}}$) is not an improvement over the previous most-preferred solution. We then continue to iteration $i+1$ by asking another comparison question to the DM: "Is the new alternative with $\underline{\mathbf{y}} = \mathbf{f}(\underline{\mathbf{x}})$ preferable to the most preferred alternative so far?" The algorithm terminates when there is no solution to FM($\mathcal{P}_i^{\preceq}$), i.e., when there are no other weights that could lead us to a challenger, indicating we have reached an optimal solution.

3.4.1.1 More than Feasibility: Middle-most Weights Model

While any solution to the feasibility model FM($\mathcal{P}_i^{\preceq}$) is a feasible weight vector, some of them may be more effective in terms of reducing the number of alternatives. Zionts and Wallenius

(1983) discussed the benefits of using the middle-most weights for faster convergence rather than picking just any feasible weights that are consistent with the DM's choices. They "maximized the slack associated with the least satisfied constraint, which is equivalent to finding the center of the largest hypersphere that can fit into the feasible region" (Zionts and Wallenius 1983, p. 526). The procedure is incorporated in later works such as Roy and Wallenius (1992) for nonlinear value functions and Stewart (1993) for discrete choice problems, and yields promising results in the rate of convergence. Therefore, we also use the following *M*iddle-most Weights Model (MWM), as an improvement over the feasibility model FM($\mathcal{P}_i^{\preceq}$):

$$
\begin{aligned}
\text{MWM}(\mathcal{P}_i^{\preceq}) \ \max_{\mathbf{w},t} \quad & t \\
\text{s.t.} \quad & \sum_{k=1}^{K} w_k y_k^n \leq \sum_{k=1}^{K} w_k y_k^p - t \qquad \forall\, (\mathbf{y^n}, \mathbf{y^p}) \in \mathcal{P}_i^{\preceq} && (3.11) \\
& \sum_{k=1}^{K} w_k = 1 && (3.12) \\
& w \geq 0 && (3.13)
\end{aligned}
$$

Any solution to the MWM($\mathcal{P}_i^{\preceq}$) is also a solution to FM($\mathcal{P}_i^{\preceq}$). The idea of the largest slack is also similar to the analytical center approach used by Toubia et al. (2004), and similar in spirit to the notion of maximizing discrepancy in Toffano et al. (2022).

3.4.1.2 Challenging the Most Preferred Solution: Challenger Model

Another issue that may arise within an iteration is that although we return the middle-most feasible weight $\underline{\mathbf{w}}$, the corresponding solution of the main model MM($\underline{\mathbf{w}}$) may coincide with the currently best alternative — even if it is not optimal — leading us to fail to find a challenger. When this occurs, we use the *Challenger Model* (CM) to complement the middle-most weights model. One contribution of our work is the addition of this challenger model; possibly this approach was not considered viable in the past when bilinear problems were difficult to solve. Our current aim is to find a new feasible weight vector $\mathbf{w} \neq \underline{\mathbf{w}}$ that is consistent with the DM's revealed preferences and that leads us to a solution that will challenge the most preferred solution so far, i.e., we will try to find an increase in the objective value. Let y^* be the most preferred solution so far and δ a small positive number. Then we formulate the Challenger Model (CM) as follows:

$$
\begin{aligned}
\mathrm{CM}(\mathcal{P}_i^{\preceq}, y^*, \delta) \quad \max_{\mathbf{w},\mathbf{x}} \quad & 0 \\
\text{s.t.} \quad & \sum_{k=1}^{K} w_k y_k^n \leq \sum_{k=1}^{K} w_k y_k^p \qquad \forall\, (\mathbf{y^n}, \mathbf{y^p}) \in \mathcal{P}_i^{\preceq} && (3.14) \\
& \sum_{k=1}^{K} w_k f_k(\mathbf{x}) \geq \sum_{k=1}^{K} w_k y_k^* + \delta && (3.15) \\
& \mathbf{x} \in \mathcal{X} && (3.16) \\
& \sum_{k=1}^{K} w_k = 1 && (3.17) \\
& w \geq 0 && (3.18)
\end{aligned}
$$

Constraint (3.15) forces the objective value of the returned solution to be at least δ greater than that of the most preferred solution so far. The left hand side of (3.15) is bi-linear. If $\mathrm{CM}(\mathcal{P}_i^{\preceq}, y^*, \delta)$ is feasible, the returned weight vector has the potential to yield a solution with a higher objective value; hence, it leads us to a *challenger*.

3.4.2 Structure of the Proposed Interactive Preference-based Optimization Method

The flow chart and pseudo-code of the algorithm can be found in Appendix B.1 and B.2, respectively. To initialize, we solve $\mathrm{MM}(\mathbf{w^1})$ and $\mathrm{MM}(\mathbf{w^2})$ for two arbitrary different initial weights $\mathbf{w^1}$ and $\mathbf{w^2}$. Those can be selected at random, as long as they result in different alternatives. The next step is to ask the DM's preferences between these two alternatives. Next, we solve $\mathrm{MWM}(\mathcal{P}_1^{\preceq})$ and obtain a feasible weight vector $\underline{\mathbf{w}}$. As we assume the DM is rational and her implicit value function is a convex combination of the objectives, $\mathrm{MWM}(\mathcal{P}_i^{\preceq})$

will always be feasible. Then we solve MM($\underline{\mathbf{w}}$) to find a new alternative $\underline{\mathbf{y}}$. If it returns a challenger, i.e., an alternative different than the two included in the previous comparison question to the DM, then we proceed to the second iteration of the algorithm, where we ask DM for her preferences between the preferred alternative so far and the challenger. The DM's response is added to ($\mathcal{P}_1^{\preceq}$) to obtain ($\mathcal{P}_2^{\preceq}$), and the algorithm continues by solving MWM($\mathcal{P}_2^{\preceq}$). The feasible weight space will be reduced each time there is an update based on the preference information revealed. If MM($\underline{\mathbf{w}}$) does not return a challenger, then we solve CM($\mathcal{P}_i^{\preceq}, y^*, \delta$) to find a different weight vector in line with the preference information that can yield a challenger, i.e., a solution which can result in a higher objective value than the most preferred solution so far. The algorithm terminates when CM($\mathcal{P}_i^{\preceq}, y^*, \delta$) becomes infeasible, indicating there is no other weight vector that can challenge the most preferred solution so far.

3.4.3 Theoretical Properties

In this section, we discuss basic optimality properties of our proposed method. Assume the true weights of the implicit value function of the DM are $\mathbf{w}^{DM}$, and $\mathbf{y}^{DM}$ is the optimal solution if the true weights were known. Then we say the solution $\mathbf{y}$ is optimal if and only if $\mathbf{w}^{DM^T}\mathbf{y}^{DM}=\mathbf{w}^{DM^T}\mathbf{y}$.

Proposition 1. *If the feasible region is a compact polyhedron and the implicit value function of the rational DM is in the form of a convex combination of the objectives, our algorithm terminates in a finite number of steps.*

Proof. Under the assumptions stated, there are finitely many extreme points that are candidates to be the optimal solution. The extreme points here are defined in terms of $\mathbf{x}$. However, optimality is defined in terms of $\mathbf{y}$. The number of extreme points in terms of $\mathbf{y}$ is no more than that in terms of $\mathbf{x}$, as $\mathbf{y=f(x)}$. Since there are finitely many choices of $\mathbf{x}$, there are finitely many choices of $\mathbf{y}$. Since we assume the DM is rational, transitivity holds and as each given preference information strictly reduces the remaining feasible region, there is no cycling between alternatives. As the number of candidates is finite, our algorithm terminates in finite number of steps. $\square$

Proposition 2. *Denote by* $\boldsymbol{y}^{DM}$ *the optimal solution if the true weights of the DM,* $\boldsymbol{w}^{DM}$*, were known, and* $\boldsymbol{y*}$ *the returned (most preferred) solution so far when encountering infeasible CM(*$\mathcal{P}_i^{\preceq}, y^*, \delta$*) for some given* $\delta > 0$*. If the feasible region is a compact polyhedron and the implicit value function of the rational DM is in the form of a convex combination of the objectives, then* $\boldsymbol{w}^{DM^T}\boldsymbol{y*} > \boldsymbol{w}^{DM^T}\boldsymbol{y}^{DM}$*-*$\delta$*. Also,* $\exists\ \delta > 0$ *such that the algorithm returns the optimal solution.*

Proof. Assume the most preferred solution $\mathbf{y*}$ is not within δ of the optimal solution, i.e., $\mathbf{w}^{DM^T}\mathbf{y}^{DM}>\mathbf{w}^{DM^T}\mathbf{y*}+\delta$. For the algorithm to terminate, CM($\mathcal{P}_i^{\preceq}, y^*, \delta$) must be infeasible, but $\mathbf{w}^{DM}$ and $\mathbf{y}^{DM}$ would be a feasible solution for CM($\mathcal{P}_i^{\preceq}, y^*, \delta$), contradicting CM($\mathcal{P}_i^{\preceq}, y^*, \delta$) being infeasible, so the algorithm would continue. Hence, if CM($\mathcal{P}_i^{\preceq}, y^*, \delta$) is infeasible, the most preferred solution, $\mathbf{y*}$, is always within a range δ of the optimal solution of the DM.

Next we show that $\exists\ \delta > 0$ such that the algorithm returns the optimal solution. Assume that the algorithm returns a $\mathbf{y*}$ that is not optimal. Then there is a feasible y' such that $\mathbf{w}^{DM^T}\ \mathbf{y}'>\mathbf{w}^{DM^T}\mathbf{y*}$, hence there also exists a sufficiently small $\delta > 0$ such that $\mathbf{w}^{DM^T}\mathbf{y}'\geq\mathbf{w}^{DM^T}\mathbf{y*}+\delta$. Then, $\mathbf{w}^{DM}$ and $\mathbf{y}'$ would be a feasible solution for CM($\mathcal{P}_i^{\preceq}, y^*, \delta$), contradicting CM($\mathcal{P}_i^{\preceq}, y^*, \delta$) being infeasible, so the algorithm would continue. $\square$

3.5 Numerical Experiments

In this section, we report on numerical experiments using calibrated real-life data on a hypothetical sustainable sourcing problem. We assess our approach in terms of the number of questions asked to the DM, the number of models solved and the total CPU time needed, consistent with the criteria used in the comparative review of multi-objective optimization methods by Aksoy et al. (1996). We first describe the sustainable sourcing problem. Then we share our findings with 3 and 4 criteria. To assess the performance of the proposed approach, we use the Zionts-Wallenius (1976) method as a benchmark. Future work can include benchmarking against several of the other methods reviewed earlier that were designed for other problem types.

3.5.1 Sustainable Sourcing Problem

An apparel brand (DM) produces a number of fashion products, each of which can only be made from certain blends of a set of raw materials. For example, a sustainable blouse can only be made from any blend of organic cotton, silk and recycled cotton. We assume the demand for each product type is deterministic. There are multiple suppliers, each with different sets of available raw materials, total costs and resulting environmental impacts in terms of the global warming potential (GWP) and water usage. Let $\mathcal{N} \equiv \{1, 2, \ldots, N\}$ be a

Parameter	Interpretation
d_n	Demand for product type n (in kilograms)
b_{ms}	Capacity of material m for supplier s (in kilograms)
a_{mn}	1 if material m is used to produce product type n; 0 otherwise
g_{ms}	Global warming potential per kilogram of material m from supplier s
h_{ms}	Water usage per kilogram of material m from supplier s
c_{ms}	Cost per kilogram of material m from supplier s

Table 3.1: Parameters used in the sustainable sourcing problem

set of products, $\mathcal{M} \equiv \{1, 2, \ldots, M\}$ be a set of materials, and $\mathcal{S} \equiv \{1, 2, \ldots, S\}$ be a set of suppliers. Decision variable $\mathbf{x}$ with $[x_{msn}]_{m\in\mathcal{M}, s\in\mathcal{S}, n\in\mathcal{N}}$ represents the quantity of material m in kilograms used from supplier s to produce product type n. We formulate the sustainable sourcing problem as follows:

$$
\begin{aligned}
\max_{\mathbf{x},\mathbf{y}} \quad & U(y_1, y_2, y_3) & & \\
\text{s.t.} \quad & y_1 = \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \sum_{n\in\mathcal{N}} c_{ms} x_{msn} & & (3.19) \\
& y_2 = \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \sum_{n\in\mathcal{N}} g_{ms} x_{msn} & & (3.20) \\
& y_3 = \sum_{m\in\mathcal{M}} \sum_{s\in\mathcal{S}} \sum_{n\in\mathcal{N}} h_{ms} x_{msn} & & (3.21) \\
& \sum_{n\in\mathcal{N}} x_{msn} \leq b_{ms} & \forall m \in \mathcal{M}, \forall s \in \mathcal{S} & \quad (3.22) \\
& \sum_{m\in M} \sum_{s\in S} x_{msn} \geq d_n & \forall n \in \mathcal{N} & \quad (3.23) \\
& \sum_{s\in S} x_{msn} \leq a_{mn} d_n & \forall m \in \mathcal{M}, \forall n \in \mathcal{N} & \quad (3.24) \\
& x_{msn} \geq 0 & \forall m \in \mathcal{M}, \forall s \in \mathcal{S}, \forall n \in \mathcal{N} & \quad (3.25)
\end{aligned}
$$

We maximize the DM's implicit value function. Constraints (3.19)-(3.21) define the objectives: cost, GWP and water usage. Constraint (3.22) is the capacity limits for each supplier for each material. Constraint (3.23) ensures the demand is satisfied for each product, while constraint (3.24) makes sure that each product type is composed of its allowed raw materials. This is a simplified sourcing problem, intended only as a basis for our numerical experiments and conversations with practitioners, not as an exact representation of an actual sourcing problem.

The DM has an implicit value function $U(\mathbf{y})$ that is a convex combination of the objectives y_1, y_2, y_3 defined by the weight vector $\mathbf{w}$. As less is better for all three objectives, $\max_{\mathbf{y}} U(\mathbf{y}) \equiv$

$\max_{\mathbf{y}} \mathbf{w}^T(-\mathbf{y}) \equiv \min_{\mathbf{y}} \mathbf{w}^T \mathbf{y}$. We can similarly formulate the bi-linear problem as follows:

$$
\begin{aligned}
\min_{\mathbf{w},\mathbf{x},\mathbf{y}} \quad & \sum_{k=1}^{3} w_k y_k & & \\
\text{s.t.} \quad & \text{constraint sets } (3.19) - (3.25) & & \\
& \sum_{k=1}^{3} w_k y_k^n \geq \sum_{k=1}^{3} w_k y_k^p & \forall\, (\mathbf{y^n}, \mathbf{y^p}) \in \mathcal{P}^{\preceq} & \qquad (3.26) \\
& \sum_{k=1}^{3} w_k = 1 & & \qquad (3.27) \\
& w \geq 0 & \forall k \in \mathcal{K} & \qquad (3.28)
\end{aligned}
$$

We adapted the optimization problems used in the algorithm accordingly. We generated realistic data based on the the Higg MSI. The Higg MSI is a quantitative tool to help decision makers consider the environmental impacts of the raw materials used in apparel and footwear products based on extensive Life-Cycle Assessment (LCA) data. It was maintained by the Sustainable Apparel Coalition (SAC), a trade organization which includes brands, manufacturers, retailers, government and non-government organizations and academics as members and aims to reduce the environmental and social impacts of the apparel industry around the world (Radhakrishnan 2015). For instance, in the Higg MSI, conventional cotton has GWP score 2.2 and water usage score 47.6, while polyester has GWP score 3.2 and water usage score 0.7 (SAC 2020). The Higg MSI also includes assessments of the data accuracy and the geographical representation of the data for each material. To generate the environmental impact data for different suppliers, we used normal distributions with means equal to the Higg MSI scores and standard deviations dependent on the accuracy and geographical representation of the data. To conduct numerical experiments, we coded our algorithm in Julia 1.6.2 and used Gurobi 9.1.2 to solve optimization problems on a dual core (Intel Core i7 2.40GHz) computer with 8 GB RAM. Gurobi can only find global optimal solutions to bi-linear constrained problems since recently, starting with Gurobi 9.0, which was introduced in 2020. We later also coded our algorithm in Python 3.10.2 with Gurobi

10.0 for the interviews with practitioners.

3.5.2 Numerical Experiments

For our numerical experiments we considered 20 suppliers, 15 raw materials, 10 product types and 3 objectives (cost, GWP and water usage). For known weights, this is a straightforward optimization problem with 3000 variables and 3460 constraints, but if the DM's value function is unknown, even a problem of this size is quite challenging. We tested our algorithm on 5 different randomly generated data sets, with 10 different true weight vectors (representing 10 different decision makers) and 6 different initializations for the weight vector, for a total of 300 scenarios. We used $\delta = 0.10^{-5}$ for each of the runs. Gurobi 9.1.2 can find a feasible solution in few seconds to the bilinear Challenger Model if one exists; if no feasible solution exists it can keep searching for a long time, so we set a time limit of one minute to terminate this step. We experimented with different time limits and values of δ and did not encounter any instances where a feasible solution was found after one minute. We generated the DM's answers to the pairwise comparison using the true weights. We only used those weights to determine the answers to the comparison questions, they were otherwise unknown to the algorithm. We use the number of questions asked as our main performance measure; we also report the CPU time and the number of optimization models solved. Table 3.2, 3.3 and 3.4 shows the results for the 5 data sets, number of questions asked, number of models solved and the total CPU time in seconds, respectively.

	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
Average:	15.60	16.87	14.78	16.87	17.73
Minimum:	5.00	4.00	3.00	5.00	4.00
Maximum:	26.00	26.00	23.00	27.00	32.00

Table 3.2: Number of Questions Asked in 5 Data Sets with 3 Criteria

	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
Average:	49.22	51.48	47.10	50.22	53.28
Minimum:	17.00	12.00	11.00	15.00	13.00
Maximum:	90.00	84.00	78.00	82.00	100.00

Table 3.3: Number of Models Solved in 5 Data Sets with 3 Criteria

	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
Average:	35.64	39.89	42.86	36.04	34.79
Minimum:	0.14	0.22	0.12	0.14	0.18
Maximum:	68.80	69.93	66.57	64.56	68.86

Table 3.4: Total CPU Time in Seconds for 5 Data Sets with 3 Criteria

Table 3.4 shows that the total CPU time is at most around a minute for each DM scenario. Almost all of this occurs in the last step, due to the one-minute time limit we set for the Challenger Model to stop searching for a feasible solution. The CPU time within each iteration before the last is minimal, and negligible compared to the actual time the DM would take to answer the pairwise comparison questions. Once the answers are provided, the algorithm can move on to the next iteration in seconds. The number of questions asked depends on the initializations. Figure 3.1 shows the average number of questions in each DM scenario for Data Set 1. Other data sets have similar results. The algorithm terminates after less questions in the first three DM scenarios. This is because in the first three scenarios, the DM's true weights are one of the corner points, e.g., w^{DM}=[1 0 0], and it is easier to find the optimal solution in such cases. When the true weights of the DM do not lie on an extreme (DM Scenarios 4-10), the algorithm needs to ask more questions before terminating.

Next, we look at how close we get to the optimal solution in each iteration, using the relative optimality gap, measured by $(\mathbf{w^*}^T\mathbf{y}_i - \mathbf{w^*}^T\mathbf{y^*})/\mathbf{w^*}^T\mathbf{y^*}$, where $\mathbf{y}_i$ is the solution at iteration i. To illustrate, the change in relative optimality gap by iteration for DM Scenario

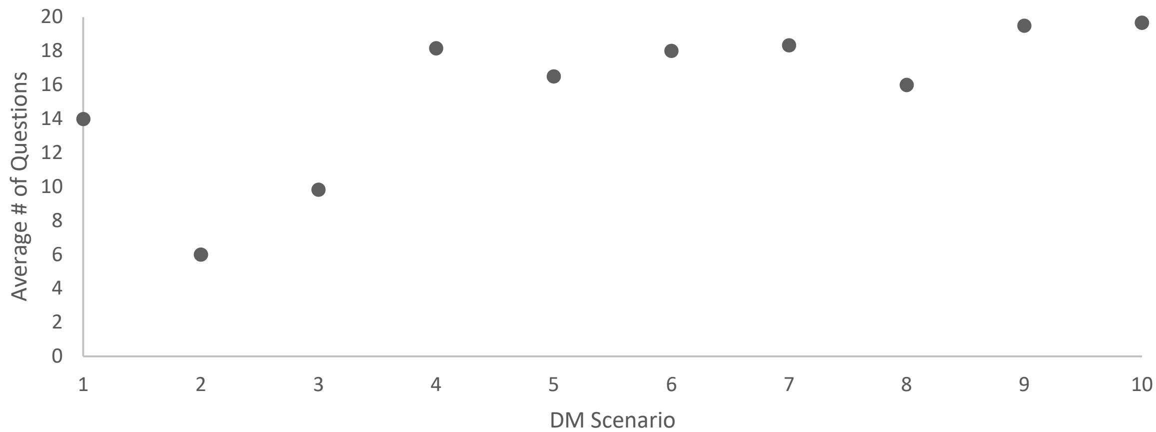


Figure 3.1: Average Number of Questions Asked versus DM Scenarios for Data Set 1

9 in Data Set 1 is illustrated in Figure 3.2, for each of the 6 different initial weight vectors. The optimality gap decreases with each iteration. The algorithm converges to the optimal solution in 10 questions or less for this case. Next, we show the distribution of the number of iterations to reach the optimal solution across all 300 scenarios, in Figure 3.3.

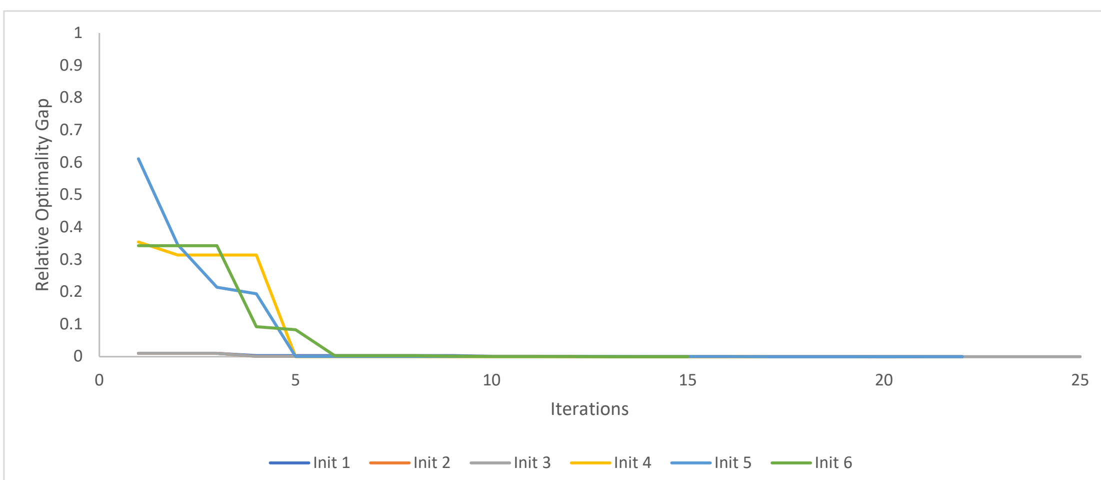
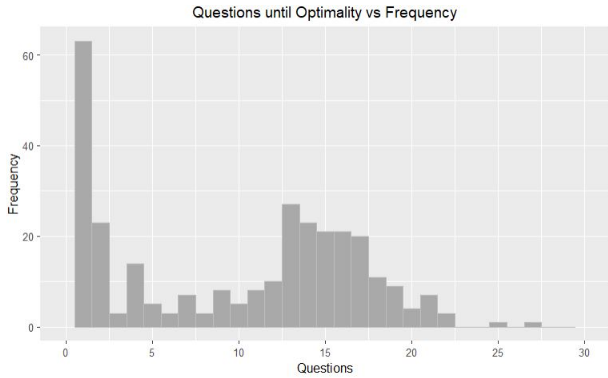
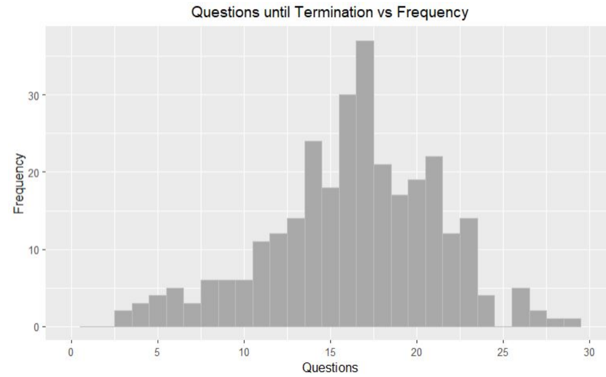


Figure 3.2: Relative Optimality Gap versus Iteration for DM Scenario 9 for Data Set 1

Figure 3.3b shows the distribution of the number of questions asked until termination. Figure 3.3a shows the number of questions needed to reach the optimal solution, although



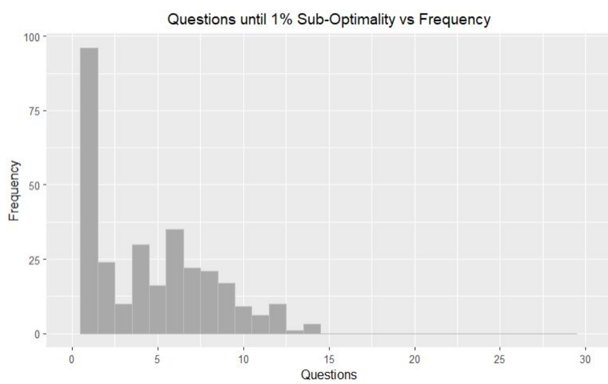
(a) The Distribution of Number of Questions (Iterations) until Optimality



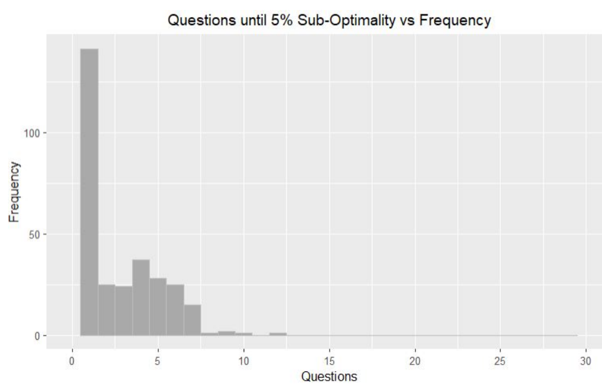
(b) The Distribution of Number of Questions (Iterations) until Termination

Figure 3.3: The Distribution of Number of Questions until Termination and Optimality

the algorithm has not necessarily confirmed optimality at that stage. The algorithm actually converges to the optimal solution faster but asks further questions to confirm (near-)optimality of the current solution. Figures 3.4a-3.4b show similar distributions for 1% and 5% optimality.



(a) The Distribution of Number of Questions (Iterations) until 1% Sub-Optimality



(b) The Distribution of Number of Questions (Iterations) until 5% Sub-Optimality

Figure 3.4: The Distribution of Number of Questions (Iterations) until 1% and 5% Sub-Optimality

Once we allow for a 1% gap relative to the optimal solution, the number of questions needed drops significantly. The maximum number of questions needed to reach optimality was 27 and allowing for 1% sub-optimality, the maximum number of questions needed drops to 14. If we relax further to 5% even less questions are required, but the additional gain is limited. Note that the uncertainty in the data in this kind of problem is likely much greater than 1% or 5%.

3.5.3 Numerical Experiments: 4 Criteria

We increased the number of criteria from 3 to 4 by adding fossil fuel usage as another environmental criterion following the Higg MSI. Table 3.5 shows the results for a random dataset with 10 different DM scenarios and 10 different initializations.

	Questions Asked	Models Solved	CPU Time (sec)
Average:	21.04	64.88	43.15
Minimum:	2.00	7.00	0.05
Maximum:	48.00	146.00	75.03

Table 3.5: Overview of the Results for 4 Criteria

On average, we terminate after 21 questions, and the average CPU time is still around a minute and it does not exceed 90 seconds, which is promising. The evolution of the relative optimality gap is illustrated in Figure 3.5. We converge to the optimal solution in 28 questions or less. The algorithm reaches 1% sub-optimality in at most 14 questions.

3.5.4 Benchmarking with the ZW Method

Here we explain the similarities and differences between our interactive optimization approach (IOA) and the classic ZW method. The aim of Zionts and Wallenius (1976) was to develop a method that can work well in practice. They assume an additive linear implicit

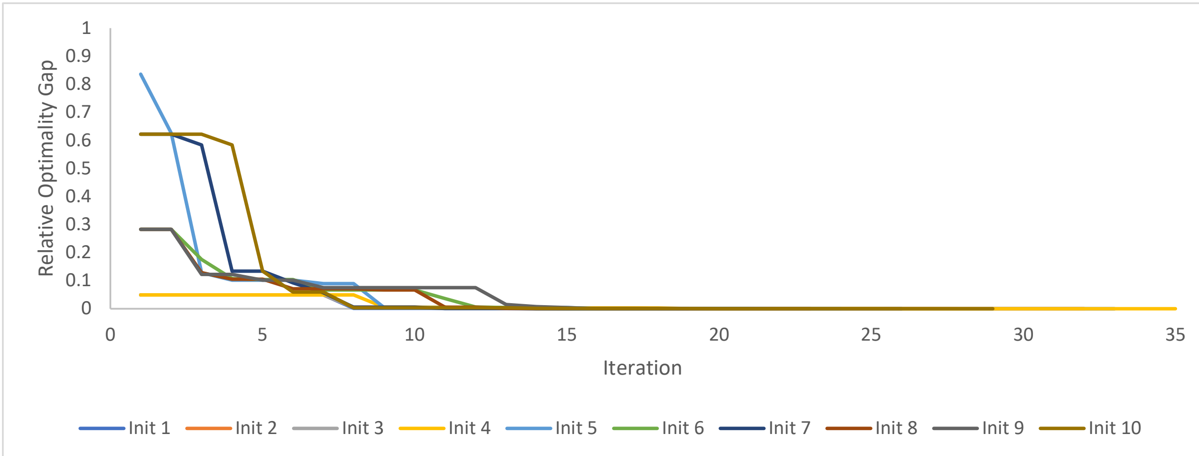


Figure 3.5: Relative Optimality Gap versus Iteration for DM Scenario 7 for 4 Criteria

utility function. Their method is similar to the simplex algorithm, but as the objective function is unknown a priori, the decision maker decides which variable should enter the basis by specifying preference information. The ZW method begins by choosing arbitrary multipliers (weights) for the objective function and finds an efficient (Pareto optimal) solution. Then the subset of efficient variables are selected among the set of nonbasic variables; variables are called efficient if they result in an efficient solution when introduced into the basis. For each efficient variable, which is adjacent to the current solution, a set of trade-offs are defined and the DM is asked whether the trade-off seems attractive or not. According to her answers, a new set of weights is generated and the process repeats. The algorithm stops when no trade-off with an adjacent efficient solution seems attractive.

The similarities between our approach and the ZW method can be summarized as follows: Our algorithm also searches efficient solutions and we also ask pairwise comparison questions, although we ask for direct orderings rather than trade-offs. We generate consistent weights in the next iteration based on the DM's answers in a similar fashion. The main difference is that we do not base our algorithm on the simplex method, i.e., we do not ask pairwise comparisons based on adjacent efficient solutions. Instead, we create a different challenger by using weights consistent with the DM's preferences and only ask a comparison question between that and the most preferred solution so far.

One of the main considerations in the performance of interactive algorithms is the number of questions asked. We initialize our algorithm with two different random weights yielding two different alternatives, which can be far away from each other. The comparison question becomes relatively easy to answer and it can result in a better cut to the feasible region compared to a more local comparison as in the ZW method.

We solved the same sustainable sourcing problem with 3 criteria (data set 1) with the ZW method using middle-most weights to assess our algorithm's performance. We were not able to find the optimal solution with the ZW method in around half of the cases as our implementation of their algorithm encountered numerical issues, presumably due to the very small differences between adjacent solutions, which appeared to clash with the numerical precision of Gurobi. The comparison of our algorithm versus the ZW method (for those cases where the optimal solution is obtained) is shown in Table 3.6, which shows that our method obtained the optimal solution with less questions than the ZW method. Moreover, our approach requires far less CPU time. Hence, we believe our interactive optimization approach is promising for sustainable decision making.

	Questions Asked		Models Solved		CPU Time (sec)	
	IPO	ZW	IPO	ZW	IPO	ZW
Average:	15.60	17.14	49.22	15866.71	35.64	402.75
Minimum:	5.00	1.00	17.00	3001.00	0.14	86.61
Maximum:	26.00	54.00	90.00	39025.00	68.80	1000.48

Table 3.6: Performance Comparison of the Proposed Interactive Preference-based Optimization (IPO) and ZW Method

3.6 Initial Reactions from Practitioners

The numerical experiments indicate that the method converges to an optimal solution with relatively few steps, and that each iteration is quick enough that the decision-maker is not kept waiting. This is promising, but we also wanted to get some initial reactions from practitioners. To do so, we conducted conversations (30-45 minutes each) over Zoom with 8 individuals with experience in sourcing, sustainability, or decision analysis. Five of them worked at well-known apparel brands, one at a major consumer goods firm, and two worked as consultants and academics focusing on procurement or decision analysis. Each conversation started with a brief explanation of the hypothetical sustainable sourcing problem studied in this chapter. The practitioner was then asked to take the role of the decision-maker, and we executed the algorithm with them, presenting them with a series of binary choices with the same three attributes as in our numerical experiments, until it converged, in less than 10 minutes. After that, we asked them for their reactions to the method and how it compared to approaches currently in use in practice, and what strengths and weaknesses they saw in the algorithm. Below we summarize some of the themes that emerged. Given the informal nature of these conversations, these reactions are obviously not intended as definitive assessments of the performance of this specific algorithm, but as initial reactions and pointers to areas for further development of interactive algorithms for multi-objective optimization in general. (The quotes below have been edited lightly for clarity and grammar.)

3.6.1 Comparison to Existing Methods in Practice

None of the participants were aware of use of formal methods for multi-objective optimization in the domains they were familiar with. One said that a typical practice currently might be to come up with a limited set of scenarios and only compare those, hoping that the comparison set included good solutions. Another observed that many current approaches rely on checkboxes and spreadsheets, and considered that a “very clunky” way to make decisions. Another responded that their approach was to drastically reduce the set of options by limiting the number of suppliers to two or three, and then manually compare between those.

A typical approach to this type of problem might be to elicit weights first, and then solve the optimization problem using those weights. Several of the practitioners confirmed that it would be difficult for them to provide weights: "I don't think I can", it is "very hard to weigh them", "I think it's hard to do". One commented that weights should also depend on the relative differences between the alternatives: if two alternatives are only 10% apart in terms of cost, cost would be weighted less than if they were 50% apart. They felt that the algorithm would probably be helpful with reconciling such cases. Another highlighted an additional challenge of eliciting weights in advance, pointing out that weights change all the time. Another recalled that in the early days of the Higg MSI there was a focus on weighting of attributes, but that the developers moved away from that approach over time. Several participants suggested that, rather than attempting to determine weights for the attributes, they rely more on approaches involving elimination of alternatives. One participant was comfortable with the notion of eliciting weights, based on their experience with some decision analysis tools in practice, but recognized that those specific tools really only work with a limited number of alternatives, maybe 6 or 7; commenting on the algorithm, they said that "a tool like this could really help you to manage the complexity of it and and get to something optimal with only asking like 14 questions in my case, which was really really helpful."

3.6.2 Potential Value of the Algorithm

All participants commented that they saw promise in the algorithm, though we recognize that conducting these conversations one-on-one over Zoom may make them less willing to be critical. More informative are their comments about which specific aspects of the algorithm they found valuable. One participant who had interacted with many procurement managers said that any tool has to be "drop-dead easy", and saw real opportunity in this approach. Several of them found the method easy to use. One said "I liked it. I think it was really easy to use." Another said "I think it's really easy to use. [...] I think it's a super easy way to do those trade offs and understand the different costs." A third said "my biggest take away as a buyer is that it was quite nice to do the one on one comparison instead of versus a benchmark".

Several respondents commented on ways that an algorithm like this could be useful in their environment. One said "I like the output, I think the tool is very useful. I feel like I can use it in my role, where I am looking at cost and the environmental impact and how to reduce it." Another observed that "current approaches do not showcase environmental impact aligned with cost impact in the way the algorithm does", believing that it is "good to take environmental impact into consideration at the beginning". Another felt that "the multidimensional criteria that you are trying to balance here really fits into the logic a procurement manager has to consider as they think through their procurement decisions." One suggested that there is a risk of getting "hemmed in" with these types of decisions, suggesting that if you tell buyers to focus on water they may take that too literally and make sourcing decisions that do not appropriately balance multiple objectives; that respondent seemed to imply that a more dynamic approach as offered by the algorithm could mitigate that.

3.6.3 Providing Insight into Decision-Maker's Value Function

Several practitioners suggested that this approach can help provide insights into the nature of the trade-offs and into firms' priorities and their own. One illustrated the insight they obtained as follows: "I like the method, you can kind of talk yourself through the decision making and understand what the options are or what the impacts of the options are." Another said that "it was great to think about the trade-offs between environmental attributes", and another went further by saying "It was nice to see what I prioritize in making decisions. It was a great challenge to think about the trade-offs, especially between environmental attributes. I didn't think much about that before. It helped in understanding my own values." One thought it could be a "very valuable exercise for a business to say how are we making the decisions on this", another said "the outcome can create clarity for a business around their values and for the teams outlining sustainability goals", and a third pointed out it could be an interesting way for a firm to reconsider their weights." The challenge for a firm to articulate their values was illustrated by one practitioner as follows: "I think businesses get spun around, not knowing what the right direction to go regarding sustainability is, so I think they're confused. But I also think every business is unique and has their own values. So this is a great system, because one, it can give you direction so you can stop spinning around and second, you could put your values and your own businesses' values and constraints in it. So I think that's a very big sell." Another said "I also believe this algorithm can be very useful for small businesses or start ups. It would also help them understand how much they value each criteria and also can guide them in what aspects to prioritize."

One respondent added another twist, by commenting that this approach can help to reduce preconceived notions: "you're not telling me what the scenarios are. So you're very objective, only looking at the impacts and you don't get derailed in your mind or biased in your mind by the predetermined notions about what is good or what is bad, or what you expect to see. You're really objectively only looking at what is the outcome. And then later, you find out that this was the scenario that is linked to that outcome, instead of looking at the scenario descriptions."

3.6.4 Future Directions and Challenges

The participants pointed to several directions to consider further, and several likely challenges adopting this type of approach would face in practice. For individuals or firms to adopt a method like this, they would need to trust it, and part of that trust would come from comparing the method to other approaches and to whatever approach the firm currently uses.

They cautioned that data are often limited, and that firms are often not willing to share the necessary data (which was one of the objectives behind the Higg MSI); they pointed out that traceability in apparel supply chains continues to be limited, especially with criteria such as forced labor. One noted that the algorithm would need to be maintained, as the parameters keep changing, especially in the fashion industry where entire assortments can change several times in a year.

Several participants pointed to ways to expand the approach, by changing the objectives or possibly adding more objectives, though they recognized the challenges that would bring. Making choices with three attributes is already hard, four or more may become too difficult. Adding social attributes would be useful, though those might be hard to quantify. Specifically in the apparel sector, one reminded us that any sourcing method needs to include minimum order quantities. One pointed to scaling as a challenge: "I think you're really on to something. [...] But, as we saw in the Higg Index, the proof is in the pudding and it's very, very difficult to get that to scale."

The practitioners saw several other potential application areas, such as choosing which improvement opportunities to focus on to make a supply chain more sustainable, and how to allocate funding to those. One suggested that material sourcing in the construction industry was a promising application area, with its trade-offs between embodied energy and water in the materials and in the use of the buildings. Another saw potential value in disruption planning in supply chains, when a firm is faced with choices such as whether to use air shipments, which products can be substituted for others, etc.

3.7 Discussion and Conclusion

In this chapter, we consider sustainable sourcing in the fashion industry with unknown environmental preferences. We proposed an interactive optimization approach to solve this problem; our algorithm can be generalized to other settings with unknown implicit value functions of multiple objectives resulting in trade-offs. We build on an existing stream of work in multi-objective optimization with implicit value functions and our method follows a similar structure to some earlier ones, but we use a different combination of optimization models to generate alternatives to present to the DM. Our initial experience with the algorithm was favorable in the sense that we were able to find the optimal solution in a reasonable number of questions, and we reach 1% sub-optimal solutions in surprisingly few iterations. That is important if we want to restrict the number of questions asked to the DM.

We conducted interviews with practitioners and asked them to solve the hypothetical sustainable sourcing problem using our method. The practitioners seemed to see promise in this type of approach. They suggested that in addition to finding the optimal solution, the method can be useful in better understanding the decision maker's and/or firm's values, which may result in improving their decision making process. They highlighted the importance of having a method that is quick and easy to use. Our algorithm terminates in around a minute, even with 4 criteria; the time it takes for the DM to answer the pairwise comparison questions is higher than the CPU time the algorithm takes. Once the DM gives her preference information, she does not have to wait long at all for the next question.

The method we propose adds value when the decision problem is continuous and there are too many alternatives to present to the DM for her to make a rational decision. If the underlying value function is known in advance with certainty, then our method is not needed, as one can just use those known weights and optimize the decision problem accordingly. However, that is rarely the case in real life and it is often difficult to elicit the weights of the DM, especially when the attributes to trade-off are less tangible, more abstract, and more emotionally charged, as with problems involving environmental or social aspects. In addition, our method can work directly with the DM without needing an analyst to guide the process, as it is fully automated.

One assumption of our model is that the DM can perfectly express her preferences in line with her underlying implicit value function. However, in practice, there are several cases in which this may not occur. Some of the reasons are not paying enough attention, getting tired, changing their minds, and a linear value function being too simple. The behavioral study by Korhonen et al. (2012) finds that linear value functions can often explain choices; they recommend accepting a number of inconsistent responses unless there are too many of them in the decision aiding strategy. To adequately incorporate inconsistency in our algorithm, more work is needed.

One natural next step would be to do broader testing of the interactive method we propose. Comparing our approach to existing methods would help us better understand its potential practical use. Other extensions include exploring how to accommodate more attributes, how to deal with non-linear and perhaps discontinuous value functions, and mixed-integer optimization problems. several of the other methods we reviewed earlier may be more applicable in some such cases. Overall we hope that this work will encourage scholars in sustainable operations to explore the field of multi-objective optimization with unknown value functions more deeply.

CHAPTER 4

Experimental Framework for Comparative Evaluation of Interactive Preference-based Optimization Using oTree

4.1 Introduction

Through our interactions with industry practitioners, as described in Chapter 3, we gain valuable insights indicating a preference among decision-makers for the interactive optimization approach rather than using the more conventional decision-making method of eliciting weights first and then taking those weights as the true weights of the decision-maker and optimizing the given problem at hand using those weights.

It is important to note that our interactions with practitioners were limited by the small sample size. This points to the necessity for a more comprehensive study to test whether the interactive optimization approach is favored over a more traditional method used to make decisions when weights are unknown. Thus, we propose conducting further more formal experiments with a significantly larger sample size, thereby enhancing the robustness and versatility of our findings. It would also help us better understand the potential practical use of the interactive optimization approach.

Moreover, we recognize the potential influence of direct communication with the industry practitioners during the decision making process on their responses regarding the interactive optimization approach, so we need an experimental framework that can be run autonomously, without intervention by the researcher. In this chapter, we provide a framework for such an automated experiment by making the interactive optimization approach run online, within an experimental environment offered by oTree, which eliminates the need for direct communication. By this automation, we can mitigate potential biases stemming from human interaction, thereby ensuring the integrity and validity of our experimental outcomes.

To compare the interactive optimization approach with a traditional weight elicitation method and then optimizing using the elicited weights, we need to select an appropriate weight elicitation method. Many approaches exist, each with pros and cons. If we choose one which is overly complex, or one which is known to perform relatively poorly, that would unfairly bias the experiments in favor of the interactive algorithm. The method we select should be intuitive and easy to use, to ensure a fair comparison, and to minimize the cognitive burden on the participants, as the experiment also includes the interactive optimization approach and a survey afterwards. After carefully reviewing the weight elicitation techniques, with the help of Weighting Methods Selection Software (Cinelli and Miebs 2022), we narrowed down the candidate methods to direct rating, point allocation and swing weighting. Based on the comments of Aubert et al. (2020) regarding the online experiment they conducted, and the results of Bottomley et al. (2000)'s experiment comparing direct rating and point allocation (more on both of which below), we decided to use direct rating as the most appropriate benchmark.

The main aim of this study is to provide an experimental framework to test whether the interactive optimization approach is generally preferred to weight elicitation using direct rating and subsequent optimization using those weights. We speculate, based on the practitioners' feedback, that the iterative and interactive nature of the interactive algorithm helps the decision-maker have more confidence that the final solution is indeed consistent with their (unknown) preferences. If so, then participants who provided more input, i.e., for who the interactive algorithm needed more iterations, might show a stronger preference for the interactive algorithm than participants for who it converged faster. To test this, we would look at whether being asked more comparison questions in the interactive optimization approach increases their confidence in the method. If this is confirmed, that would provide a contrary perspective to the more typical desire in optimization to minmize the number of iterations.

In addition, we expose participants to both methods, and will investigate whether doing one decision-making method first versus last has any effect on the overall preference for that method. Participants who use the interactive optimization approach last may develop a greater preference for it over time, as they become more familiar with the concept through continued participation. On the other hand, participants may experience fatigue towards the end of the survey and, as a result, those who use the interactive optimization method initially may exhibit a stronger preference for it compared to those who use it later. Participants who use the interactive algorithm last may appreciate it more, having first been confronted with the challenge of providing weights using direct rating, or they may appreciate it less if they consider it too complicated relative to direct rating. To test for sequencing effects, we randomize the order of the decision-making methods participants use.

Furthermore, sequencing effects are well-documented in many other settings. It is conceivable that the sequence of attributes (cost, global warming, water use) shown to the participants has an effect on the final outcome they choose, if they inadvertently place a higher weight on whichever attribute is listed first. Because the interactive algorithm requires more steps, each involving all three attributes, we speculate that it is less sensitive to such sequencing effects than direct rating, in which the participants are only confronted with the attributes once. To explore this, we compare two treatments, in which the sequence of attributes are either cost, global warming potential and water use, or water use, global warming potential and cost, respectively.

We use oTree to make the interactive optimization approach online and design the experiment by integrating Gurobi with the Web Licence Service for solving the optimization problems. We then use Heroku as the database and create a Prolific study to recruit participants.

The outline of this chapter is as follows. In Section 2, we will provide a brief literature review and background of our study. In Section 3, we will discuss our research questions and in Section 4, we will outline the design and implementation framework of the experiments. In Section 5, we summarize and conclude our experimental framework.

4.2 Background and Literature Review

There is an extensive literature that compares different multi criteria decision making methods. Many of those studies compare the performance of the multi criteria decision making methods by conducting numerical experiments or empirical analyses by analyzing performance on existing datasets. Some examples include Zanakis et al. (1998), Mousseau et al. (2001), Ceballos et al. (2016) and Wu et al. (2023).

In addition, several studies compare the performance of different multi criteria decision making methods by conducting experiments. However, majority of those studies compare methods that can be used for multi attribute decision making. For example, Stillwell et al. (1983) conduct an experiment to compare six different multi attribute decision making methods in the context of evaluating credit applications. Hobbs and Meier (1994) compare direct weight assessment, trade-off weight assessment, additive value functions, and goal programming in an experiment to choose a resource portfolio for Seattle City Light on planners and interest group representatives. Wang and Yang (1998) evaluate the performance of three multi attribute weight measurement methods in terms of theoretical validity, predictive performance, and perceived performance in an experimental study. Hodgett (2016) conduct an experimental study to assess three multi attribute decision making methods on a technology manager and a team of nine people in an imaging company for an equipment selection problem. Ishizaka and Siraj (2018) assess the usefulness of three multi attribute decision making tools using incentive-based experiments. In this chapter, we develop an experimental framework to compare the interactive optimization approach, which is a method aimed to be used in multi objective decision making, with a weight elicitation method followed by optimization taking those weights as the true weights of the decision maker.

Selecting which method to use in multi criteria decision making is not trivial and Cinelli et al. (2022) propose a new methodology and decision support system to recommend multi criteria decision analysis methods. To decide which weight elicitation technique to use, we benefited from the Weighting Methods Selection Software (Cinelli and Miebs 2022) and narrowed down the candidate methods to direct rating, point allocation and swing weighting.

Aubert et al. (2020) conducted an online experiment on the use of swing weighting in environmental decision making. They emphasize the need for more online work and real life testing that focuses on methods like swing weighting and environmental attributes. The problem that they focus on is future wastewater infrastructure, and they recruited participants that are Paris citizens through a survey company. 298 of them used swing weighting while 357 of them used direct rating and the aim of the study was to test whether participants learn facts about wastewater management by eliciting swing weights, whether participants were able to follow the process instructions of the online survey to elicit swing weights and whether they learn about their preferences. They found evidence for limited learning about the context. Moreover, very few participants complied with the swing weighting instructions, so the authors added an exploratory fourth research question to test whether participants who did not comply with the process to elicit swing weights performed a direct rating of objectives. They found that the participants who did not follow the instructions were somewhere in between direct rating and swing weighting. Based on their findings, we decided not to pursue swing weighting as the weight elicitation process in our experimental framework.

Bottomley et al. (2000) compared direct rating and point allocation by conducting an experimental study to test several hypotheses: the weights generated by direct rating will be more reliable than of point allocation; experts, the ones who are more knowledgeable on the topic/product, will be more consistent between test and retest; direct rating will be more reliable than point allocation in the test and retest procedure; and participants will favor direct rating over point allocation. They recruited 113 undergraduate and post graduate business school students to conduct the experiment, which was about a car selection among different alternatives that are evaluated on nine attributes. They found that direct rating was more reliable over test-retest than point allocation and found medium support for experts being more consistent with direct rating. To assess whether direct rating would yield more consistent choices, they calculated the weights for both methods and tested whether the participants would end up with the same alternative in test-retest. They found that direct rating produced more consistent results and the participants also favored direct rating over point allocation. In light of these findings, we select direct rating as the weight elicitation method to use in the experimental framework we present.

In addition to just comparing the interactive optimization approach with direct rating, we would like to test what might affect this preference. In particular, we would like to observe whether the ordering of which method the participants use first affects their preferred method. Carlsson et al. (2012) conduct an experimental study on ordering effects in choice experiments, and particularly how participants' preferences are affected by the learning processes in a sequence of choice sets. They found evidence of changes in preferences and a reduction in the error variance for the last choice sets compared to the first ones. Similarly, the framing of the questions in terms of the attributes shown may influence the final outcome. Oppewal et al. (2015) compares two choice experiments on how destination and experience information affect holiday choices that differ in how the attribute information is presented. Their findings show that presenting an attribute early in the task enhances its importance. To test for both ordering effects above on the preference of the decision making method, we outline four treatments, in which we change the order of the decision making method and the sequence of the attributes shown.

A crucial step in our experimental framework is to adapt the interactive optimization approach for online use. To facilitate this, we have decided to utilize oTree, which is a web-based tool designed for conducting interactive experiments in various settings such as laboratories, online platforms, or field studies, either individually or in combination (Chen et al. 2016). It operates directly through a web browser, so the participants do not need to install any software on their devices and it is compatible with any device with internet connectivity, including desktop computers, tablets, and smartphones. It is a widely used tool in the economics literature, especially behavioral economics and decision analysis. Some recent examples include Choi et al. (2022), Laudenbach et al. (2023), Ifcher and Zarghamee (2023) and Drobner and Goerg (2024) . Even though not common in operations management, more studies are emerging in the field of operations management that utilize oTree. To name a few, Davis et al. (2022) conduct experiments to explore different inventory-sharing methods within a two-tier supply chain comprising an upstream manufacturer and two downstream retailers and Walker et al. (2023) create a novel game structure and experimental framework to provide managerial insights into how principles of retainage mediate trust and trustworthiness in competitive procurement environments dealing with moral hazard. The integration of optimization that requires an optimization software with experiments in oTree is not commonly observed in literature. One of the contributions of our study to the literature is to illustrate the possibility of this integration.

4.3 Research Questions

In this section, we will describe our research questions and briefly discuss what various hypothetical results would imply.

4.3.1 Preference for the Interactive Optimization Approach

The main goal of our experiments is to compare the interactive optimization approach with a more traditional method, direct rating followed by optimization with known weights. From the interactions with the industry practitioners in Chapter 3, we have observed that decision-makers appear to prefer interactive optimization approach over weight elicitation by direct rating. However, the sample of practitioners was small and they did not actually do the direct rating method itself. In addition, having the researcher directly communicate with a participant might bias their responses. Thus, it would be beneficial to test Research Question 1 with a much larger sample, where the experimental setting is automated without the need of communicating with anyone. Hence, our first research question can be outlined as follows:

Research Question 1. *Do decision-makers generally prefer the interactive optimization approach over direct rating followed by optimization?*

If the decision-makers generally prefer interactive optimization approach over direct rating followed by optimization, then the results would be consistent with our interactions with practitioners. This finding would encourage the integration of interactive optimization methodologies within business models and decision-making processes. Furthermore, this shift could also promote more research to have a deeper understanding of the practical feasibility and implications of integrating these methodologies into real-world contexts. Field experiments, in particular, would play a crucial role in identifying implementation challenges and strategies to mitigate them effectively.

If the decision-makers generally prefer direct rating over interactive optimization approach, then it would signify that decision-makers prefer more conventional strategies. The reason behind such preference might be a lack of understanding of the mechanism behind the interactive optimization approach and hence a lack of trust in using it. One way to mitigate this issue would be to work on how to convey the logic and the mechanism of the interactive optimization approach in a better way so that decision-makers will be more comfortable adopting it in their decision-making processes.

4.3.2 Consistency between Method Preference and Choosing the Respective Optimal Solution

Given that the decision-maker's weights are not known, there is no unambiguous way to compare two different decision approaches. It is intuitive to assume that if participants give higher ratings to one approach, then they would also prefer the optimal solution generated by it. However, that is not automatically true; see for instance Beaudrie et al. (2021) who report conflicting responses from participants about which of several multi-criteria decision methods they prefer and how much the resulting outcome reflected their values. Hence, we would like to test the following research question next:

Research Question 2. *When participants give a higher overall rating to interactive optimization approach than direct rating, do they also prefer the optimal solution found by the interactive optimization approach to the one found as a result of the direct rating?*

If the findings indicate that participants give a higher overall rating to interactive optimization approach than direct rating but do not prefer the optimal solution found by it, such a contrary finding would suggest that the "best" method does not necessariliy produce the "best" outcome. That would raise the question of under which conditions this occurs more often and what might be strategies to mitigate this.

4.3.3 Number of Iterations and Confidence in the Method

Another important aspect is the effect of the number of iterations, i.e., the number of questions asked to the decision-maker during the interactive optimization approach on the participants' confidence in the method. To this end, we put forward the following research question:

Research Question 3. *When the number of iterations of the interactive optimization approach increases, does decision-makers' confidence in the interactive optimization approach also increase?*

We speculate that once participants are more involved in the decision-making process, by being asked more questions in this case, it will increase their confidence in the decision-making method, at least up to a limit. They will understand the process and the underlying mechanism better, and will feel that they have had greater input in generating the optimal solution. They will also become more familiar with the context as well. Further research could investigate for how many iterations the confidence continues to increase, and when that effect flattens out or possibly even is reversed after too many iterations.

If the increase in the number of iterations of the interactive optimization approach makes decision-makers less confident in the approach itself, then it could signify that the participants may believe there are faster ways to achieve the optimal solution. For instance, they might also be willing to accept either solution presented at the last iterations and feel that the questions asked at the end are not necessary. Another reason behind a decrease in the confidence with the increase in the number of iterations could be that towards the last iterations, the solutions are getting closer to each other and it would be more difficult for the decision-maker to make a choice. Decision-makers might feel that their decisions are not as strong and feel less confident in the solution proposed at the end.

4.3.4 Method Ordering Effect

In addition to just comparing the interactive optimization approach and direct rating, we would also like to test the effect of the order of which method is used first. There might be a difference in the overall preference of interactive optimization depending on whether it is done first or last. Therefore, we form our next research question as follows:

Research Question 4. *Is there is a difference in the overall preference for the interactive optimization approach between when participants use interactive optimization approach first and when they use it last?*

If we do find statistically significant evidence that there is a difference in the overall preference for the interactive optimization approach depending on when the participants use the method, then it would signify the existence of an ordering effect. In that case, it would be interesting to look at the direction of the difference in the overall preference of the interactive approach when participants use it first versus last. Participants might get tired towards the end and due to this fatigue effect, the participants who use interactive optimization approach first might prefer it more compared to the participants who used it at the last. On the other hand, we might also see that the participants who use interactive optimization approach last prefer it more compared to the ones who use it first, because as they continue doing the experiment, they understand the concept better and might prefer whichever method they use last.

If we do not observe such an ordering effect, that might indicate that the overall preference for the interactive optimization approach is strong compared to direct rating. In this case, it might be beneficial to use another decision-making method as benchmark to test whether ordering effects might occur in other comparisons.

4.3.5 Attribute Ordering Effect

Our final research question is about whether the sequence of attributes shown to the participants makes a difference. There is a large literature on ordering effects, including Muthulingam et al. (2013) who find strong order effects in the context of adopting energy efficiency recommendations. In our context, with unknown weights, it is conceivable that attributes that are listed first end up being weighed more heavily than if they were listed last. Further, such a sequencing effect could be stronger when participants only interact with the attributes once, as in direct rating followed by optimization, than when participants see all the attributes repeatedly, as in the interactive algorithm. This leads to the research question below.

Research Question 5. *Will decision-makers' optimal choice have a higher cost (lower* *weight on cost), when cost is shown as the last attribute rather than the first? Is any such sequencing effect weaker for the interactive algorithm than for direct rating followed by optimization?*

We believe seeing cost attribute first will have an effect on people to put more emphasis on the cost rather than the environmental attributes. Hence, we believe the final optimal solution chosen in this case will have lower average cost, compared to the case where cost was shown as the last attribute. This finding would be important as it could have the implication that analysts can nudge people into making an environmentally friendly choice by changing the ordering of the attributes by putting the non-environmental attributes last. This would immediately raise the question whether all situations involving optimization with unknown weights are sensitive to the sequence in which the attributes are listed. This could point to another benefit of interactive methods in general.

If we do not observe such ordering effects, it would be worth investigating under which conditions the ordering effect will occur. For example, if the two solutions compared are very close to each other, this effect might be more prevalent. Similarly, if the two alternatives in the pairwise comparison question are very far away from each other, it could reduce the ordering effect.

4.4 Design and Implementation Framework of Experiments

The goal of this work is to assess the reactions to the interactive optimization approach and compare it to a more traditional decision-making method, direct rating. We also would like to test whether the sequence of attributes shown to the participants change the final outcome. The sequence of the attributes shown changes from cost, global warming potential, water use to water use, global warming potential and cost, respectively. We also change the sequence of decision-making methods that participants will use. For the context, we use the same sustainable sourcing setting as Chapter 3 with the same parameters to be consistent with the numerical experiments as well as the interactions with the industry practitioners.

As we would like to test for both the ordering effect on which method is used first and the ordering effect of the attributes, we have a 2x2 design shown in Table 4.1.

	First DR then IA	First IA then DR
Cost First	Treatment 1	Treatment 3
Cost Last	Treatment 2	Treatment 4

Table 4.1: Experimental Setting with 4 Treatments

In treatment 1, the participants first use direct rating, and then use the interactive optimization approach. The sequence of the attributes shown is cost, global warming potential and water use, respectively. In treatment 2, again, the participants first use direct rating, and then use the interactive optimization approach, but this time cost is the last attribute shown.

In treatment 3, the sequence of the attributes is the same as treatment 1, i.e., the sequence of the attributes shown is cost, global warming potential and water use. This time, the order of which method is used changes. Participants first use the interactive optimization approach, and then they use direct rating. In treatment 4, compared to treatment 3, the sequence of the attributes shown is water use, global warming potential and cost, respectively.

The flow of the experiment can be outlined as follows: First, participants see the consent page. After agreeing to continue, they are given a brief introduction explaining the sustainable sourcing in the apparel industry and what they need to do in the experiment. Then they will use both decision-making methods. Depending on which treatment they see, the sequence of the methods they use varies as explained above. After experiencing both methods, optimal solution reached with both methods will be displayed to the participants and they are asked to indicate their preferences between those two solutions. The experiment will conclude with a survey at the end to collect information on their reactions to both methods and on their demographics. More detailed experimental flow of treatment 1 is outlined in Appendix C.

We incorporate three attention checks into our experiment. The first two attention check questions are asking participants a pairwise comparison of two alternatives evaluated on the three attributes used in the experiment (cost, global warming potential and water use). In both of those comparison questions, one alternative is clearly dominated by the other; so we ask the participants to select the one that performs better. The reason behind it is also to understand whether they understand how to evaluate and compare two alternatives in a pairwise comparison question. The first of those attention check questions is asked after giving the instructions and before trying any of the decision-making methods, while the second one is asked while transitioning from one decision-making method to the other one. The third attention check question is incorporated in the survey at the end, which was just asking participants to select a certain option. We also include some demographic questions about factors such as age, gender, country, education, income and environmental attitude at the end of the experiment. Each pairwise comparison question in the experiment was shown both as a table and a bar chart side by side. An example of a comparison question in the experiment is illustrated in Figure 4.1.

Title	Alternative 1	Alternative 2
Cost	2447.15	2608.25
Global Warming Potential	667.53	371.55
Water Use	483.85	610.47

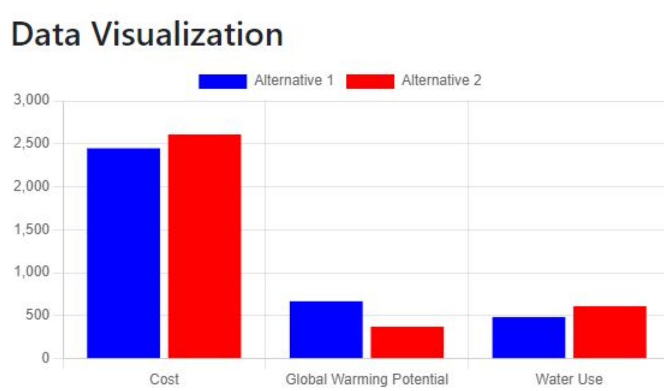


Figure 4.1: Example Pairwise Comparison from the Experiment

As aforementioned, we have decided to utilize oTree to make the interactive optimization approach online and to design our experiments. To utilize oTree, first we need to transform the code of the interactive optimization approach from Julia to Python, as Python is compatible with oTree. After transforming the code into Python, we create a web container environment and used Web License Service of Gurobi to solve the optimization problems in the experiment. We incorporate all the treatments into oTree. We also need to set up a server to run the experiment. To this end, we create an application in Heroku and deploy the oTree experiment there. With the application on Heroku using oTree, we are able to create our experiment setting. After implementing the experiment using oTree, Gurobi and Heroku, we create a Prolific study to recruit participants for our experiment. We would like to aim for a sample size of 400, that will be randomly assigned to one of the four treatments described above so that we can have approximately 100 participants per treatment. Figure 4.2 illustrates the flow chart of the experiment design process.

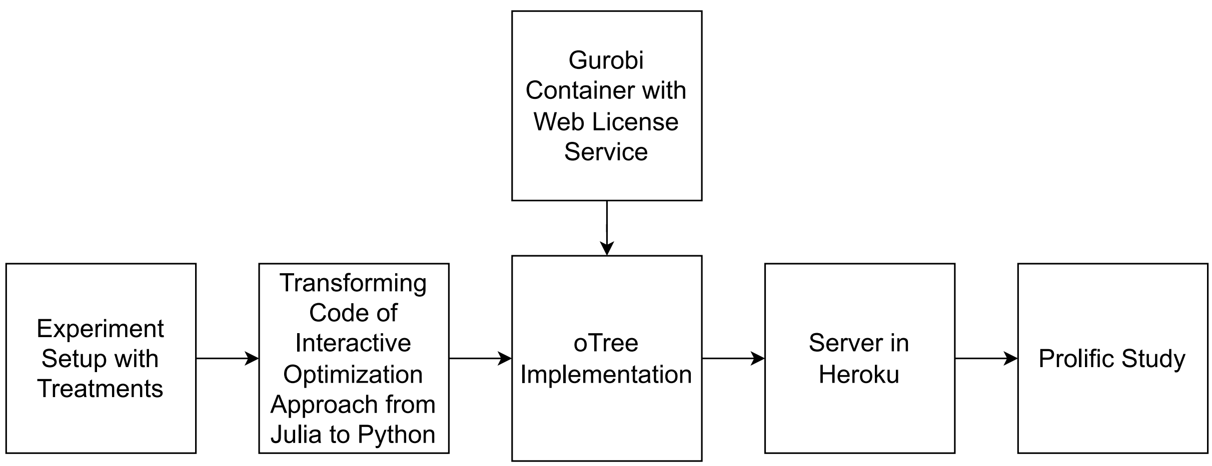


Figure 4.2: Flow Chart of the Experiment Implementation Process

The study was submitted to UCLA's Institutional Review Board with IRB#24-000537 and was certified as exempt from review on April 9, 2024.

4.5 Conclusion

In this chapter, we presented an experimental framework aimed at evaluating the general preference for the interactive optimization approach compared to weight elicitation via direct rating, followed by optimization based on those weights. Additionally, we aim to investigate whether prolonged engagement in the decision-making process enhances participants' confidence in the chosen method.

We would also like to explore the impact of the ordering of decision-making methods on overall preference. We believe that participants using the interactive optimization approach later in the sequence may develop a stronger preference for it over time, potentially due to increased familiarity. On the other hand, participants may exhibit fatigue towards the end of the survey, potentially influencing their preference. To test this ordering effect, we outline treatments, in which the sequence of the decision-making methods would change.

Furthermore, we aim to investigate whether the sequence of attributes presented to decision-makers influences the final outcome they select. The analysts can take advantage of the observation of such attribute ordering effects, as they could nudge decision-makers to make more environmentally decisions by putting the attributes that are not related to environment last. We include treatments where the sequence of attributes would change.

To implement these experiments, we utilize oTree to make the interactive optimization approach online and integrate Gurobi with the Web License Service to solve the optimization problems. Additionally, we leverage Heroku as the database platform and design a Prolific study to recruit participants. Through these experiments, we aim to gain valuable insights into decision-making processes and the factors influencing participants' preferences. Our work also highlights the integration of conducting experiments using oTree with optimization using Gurobi.

CHAPTER 5

Conclusion

With the advent of life-cycle assessment and other methods, decision-makers have increasingly good information available about the environmental consequences of their choices. However, the underlying mechanism of the decision-making process and how such choices are influenced are not known. In addition to the several tools that exist to collect environmental information, there is minimal guidance on making decisions based on such information in this inherently multi-criteria setting.

In the second chapter of this dissertation, we conduct a series of experiments and show that decision-makers are equally vulnerable to context effects when facing choices involving trade-offs between environmental attributes as they are with conventional attributes. When analysts present the environmental impacts of various alternatives to policymakers or other decision-makers, seemingly innocuous choices in how to present that information can have a major effect on which alternative is chosen. Our results highlight the importance of incorporating behavioral science into the environmental literature.

In the third chapter, we propose an interactive optimization-based approach that aims to help decision makers with difficult trade-offs and make better decisions. Our approach asks pairwise comparison questions to the decision-maker to determine the alternative most aligned with decision-maker's preferences. We test our approach on sustainable sourcing problem and use data based on Sustainable Apparel Coalition's Higg Material Sustainability Index. We reach optimality in all numerical trials with different hypothetical decision-makers. In addition to the numerical experiments, we also conduct interviews with industry practitioners to get their feedback on the proposed interactive optimization-based approach and receive very positive responses.

In the fourth chapter, we provide an experimental framework using oTree that will test the interactive optimization-based approach we developed in Chapter 3 and compare it with a more traditional approach called direct rating, followed by optimization. We also elaborate on the effect of the sequence of the attributes displayed on the final outcome as well as the effect of the sequence of the decision-making method used.

This dissertation contributes to the literature by showing that contrary to the literature broadly related to LCA, decision-makers are equally at risk of falling prey to context effects (such as attraction and compromise effects) when facing choices involving trade-offs between environmental attributes as they are with conventional attributes. Furthermore, the proposed interactive optimization approach is a valuable tool for decision-making in sustainable operations, as demonstrated by promising numerical experiments and discussions with industry practitioners. The experimental framework we present illustrates and highlights the potential of integrating Gurobi optimization in online expeirments using oTree. In general, this dissertation adds to the existing body of knowledge regarding decision-making in sustainable operations by combining behavioral and optimization-based perspectives.

APPENDIX A

Appendix to Chapter 2

This appendix provides background on the following topics regarding chapter 2:

1. The Higg Materials Sustainability Index (MSI)
2. How we transformed the choice sets in Simonson (1989)'s experiments to choices environmental attributes
3. Descriptive statistics of the different samples
4. Graphs showing initial results for all experiments
5. Details on the statistical analysis using logistic regression
6. The scales used for environmental knowledge and concern
7. Brief discussion of limitations of the study

A.1 Background on the Higg MSI (Materials Sustainability Index)

The Higg Materials Sustainability Index (MSI) is an increasingly widely used tool in the apparel industry. It provides users with quantitative information on the environmental impact of a wide range of materials used in the apparel sector in four impact categories: climate change, eutrophication, abiotic resource depletion / fossil fuels, and water resources depletion / scarcity. The index uses information from LCA studies and from industry sources and allows users to perform comparisons between materials with standardized and verified information. See SAC (2020) for more on their methodology. Until recently managed by the Sustainable Apparel Coalition, the Higg Index is "a suite of tools for the standardized measurement of value chain sustainability, and it is central to the SAC's mission to transform businesses for exponential impact" . The Higg Materials Sustainability Index is one of several tools that form part of the suite. The index is now managed by Higg, which was launched as a public-benefit company in 2019. See Radhakrishnan (2015) for more background and history. Luo et al. (2021) include the Higg Index as one of the four main methods to evaluate environmental sustainability of textiles (alongside LCA, environmental footprint, and eco-efficiency), due to its "high potential in the commercial setting". As it becomes more established, the Higg Index is also attracting more scrutiny from the media and regulators; however, any limitations it may exhibit do not affect our findings, as our work does not depend on the actual data embedded in the tool.

A.2 Transforming Simonson (1989)'s Study: Choice Sets with Environmental Attributes

We constructed hypothetical fabrics and replaced the attributes in each of Simonson (1989)'s experiments with a pair of environmental attributes. In doing so, we kept the relative proportions of the two alternatives the same as in Simonson (1989)'s original experiments, but we adjusted the scales so that the mean environmental impact score among the alternatives was 100, with higher scores representing greater impacts. Table A.1 below shows the values from the Simonson (1989)'s original experiments and Table A.2 illustrates an example of the transformation.

Original Experiment from Simonson		
Beer	Price	Quality
A	1.9	65
B	2.8	75
C	3.1	75
D	2.2	65
Avg.	2.5	70

Table A.1: The Values Used in the Beer Experiment of Simonson (1989)

In Simonson's original beer experiment choice set, there were 4 alternatives in total, with the price and quality combinations shown above. We used the same experiment but changed its name to "soda" to be more inclusive. Note that a higher quality score means a better alternative.

Rearranged for "lower is better"			Rescaled for "average = 100"		
Option	Attribute 1	Attribute 2	T-shirt	Water scarcity	Eutrophication
A	1.9	75	A	76	107
B	2.8	65	B	112	93
C	3.1	65	C	124	93
D	2.2	75	D	88	107
Avg.	2.5	70	Avg.	100	100

Table A.2: The Transformation of the Values to be Used in our Experiments

For environmental impact, a lower score means a better alternative, so we first transformed the quality values (Attribute 2) so that a lower score would correspond to a better alternative in that attribute (middle panel above). Finally, we re-scaled the numbers so that their average would be 100. The average is taken across all alternatives in the experiment; individual participants did not see all alternatives so the average score that any individual participant saw could deviate from 100.

A.3 Descriptive Statistics

The data and code are available upon request. In our analyses we only include the complete responses that pass the attention check questions. The samples sizes and selected descriptive statistics for each subsample (Behavioral Lab (BLab), Amazon Mechanical Turk (MTurk), Sustainability Institute Students (SustIS), Sustainability Institute Alumni (SustIA)) and for the combined sample are illustrated below in Tables A.3 - A.12.

Moderator Name	Min	Max	Mean	Std. Dev.
age (years)	18	34	19.93	1.36
environmental knowledge (scale 1-7; higher is more knowledgeable)	1	7	3.97	1.19
environmental attitude (scale 1-7; higher is more concerned)	4	7	6.09	0.67

Table A.3: Descriptive statistics of continuous moderators for the BLab sample (n=230)

Gender		Ethnicity	
Female	172	American Indian or Alaska Native	2
Male	53	Asian	128
Other	4	Black or African American	5
Prefer not to say	1	Multiple	28
		Other	15
		White	52

Table A.4: Descriptive statistics of categorical moderators for the BLab sample (n=230)

Moderator Name	Min	Max	Mean	Std. Dev.
age (years)	18	75	39.65	12.19
environmental knowledge (scale 1-7; higher is more knowledgeable)	1	6.8	3.68	1.29
environmental attitude (scale 1-7; higher is more concerned)	2.9	7	5.98	0.92

Table A.5: Descriptive statistics of continuous moderators for the MTurk sample (n=260)

Gender		Ethnicity	
Female	103	American Indian or Alaska Native	2
Male	156	Asian	12
Other	0	Black or African American	20
Prefer not to say	1	Multiple	12
		Other	5
		White	209

Table A.6: Descriptive statistics of categorical moderators for the MTurk sample (n=260)

Moderator Name	Min	Max	Mean	Std. Dev.
age (years)	16	42	24.56	4.98
environmental knowledge (scale 1-7; higher is more knowledgeable)	1.4	6.8	5.33	0.94
environmental attitude (scale 1-7; higher is more concerned)	3.6	7	6.3	0.64

Table A.7: Descriptive statistics of continuous moderators for the SustIS sample (n=148)

Gender		Ethnicity	
Female	105	American Indian or Alaska Native	1
Male	43	Asian	60
Other	0	Black or African American	3
Prefer not to say	0	Multiple	11
		Other	14
		White	59

Table A.8: Descriptive statistics of categorical moderators for the SustIS sample (n=148)

Moderator Name	Min	Max	Mean	Std. Dev.
age (years)	24	78	43	19.5
environmental knowledge (scale 1-7; higher is more knowledgeable)	3.6	7	5.61	0.98
environmental attitude (scale 1-7; higher is more concerned)	5.7	7	6.5	0.42

Table A.9: Descriptive statistics of continuous moderators for the SustIA sample (n=15)

Gender		Ethnicity	
Female	4	American Indian or Alaska Native	0
Male	9	Asian	5
Other	0	Black or African American	0
Prefer not to say	2	Multiple	0
		Other	3
		White	7

Table A.10: Descriptive statistics of categorical moderators for the SustIA sample (n=15)

Moderator Name	Min	Max	Mean	Std. Dev.
age (years)	16	78	29.36	12.51
environmental knowledge (scale 1-7; higher is more knowledgeable)	1	7	4.2	1.36
environmental attitude (scale 1-7; higher is more concerned)	2.9	7	6.1	0.78

Table A.11: Descriptive statistics of continuous moderators for all the samples combined (n=653)

Gender		Ethnicity	
Female	384	American Indian or Alaska Native	5
Male	261	Asian	205
Other	4	Black or African American	28
Prefer not to say	4	Multiple	51
		Other	37
		White	327

Table A.12: Descriptive statistics of categorical moderators for all the samples combined (n=653)

A.4 Initial Analyses

We first conducted preliminary analyses by visual inspection of the data looking at the choice proportions of each alternative for all samples and choice questions to determine if we observe any patterns consistent with attraction or compromise effects. The figures below represent the visual results for each pair of experiments with their corresponding choice proportions.

A.4.1 Attraction Effect

The Figures A.1 - A.4 below illustrate the results for each pair of experiments for the attraction effect with their corresponding choice proportions.

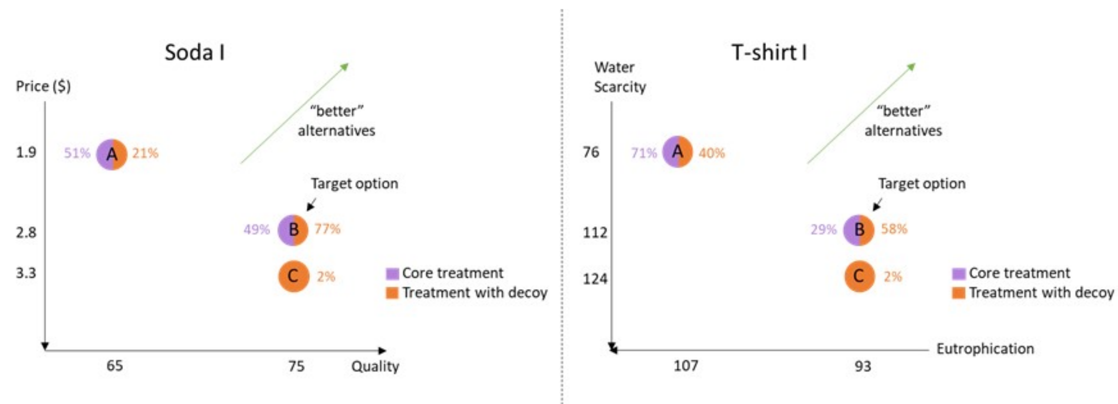


Figure A.1: Absolute choice proportions for the Soda I and T-shirt I experiments

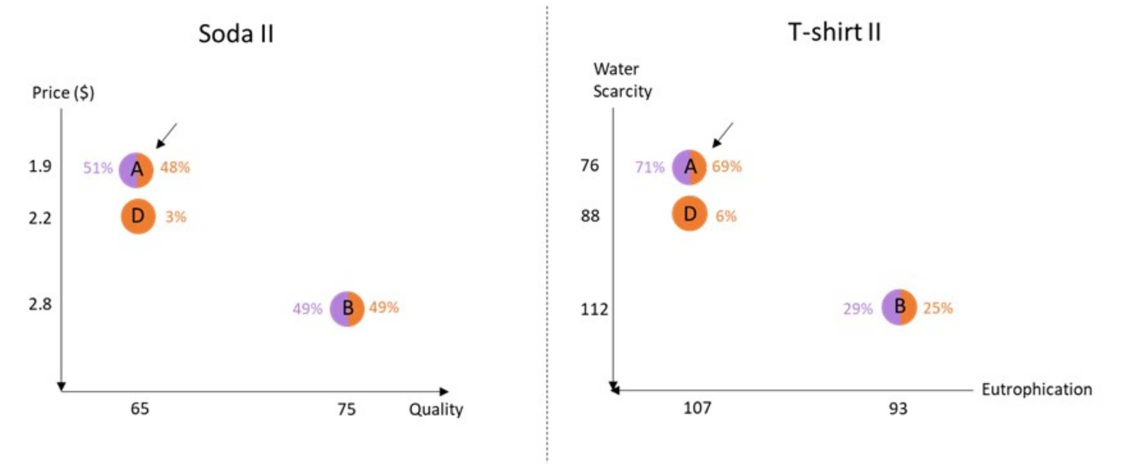


Figure A.2: Absolute choice proportions for the Soda II and T-shirt II experiments

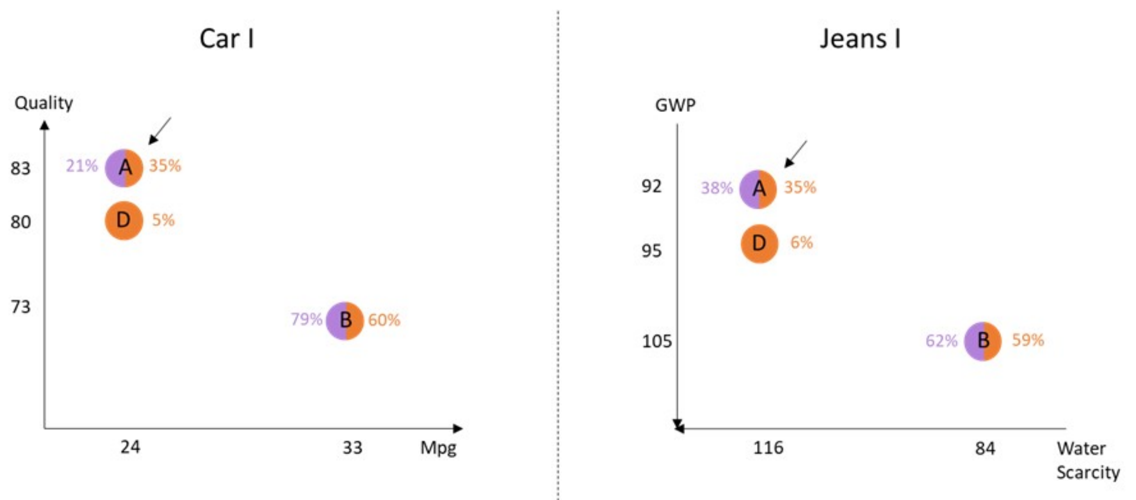


Figure A.3: Absolute choice proportions for the Car I and Jeans I experiments

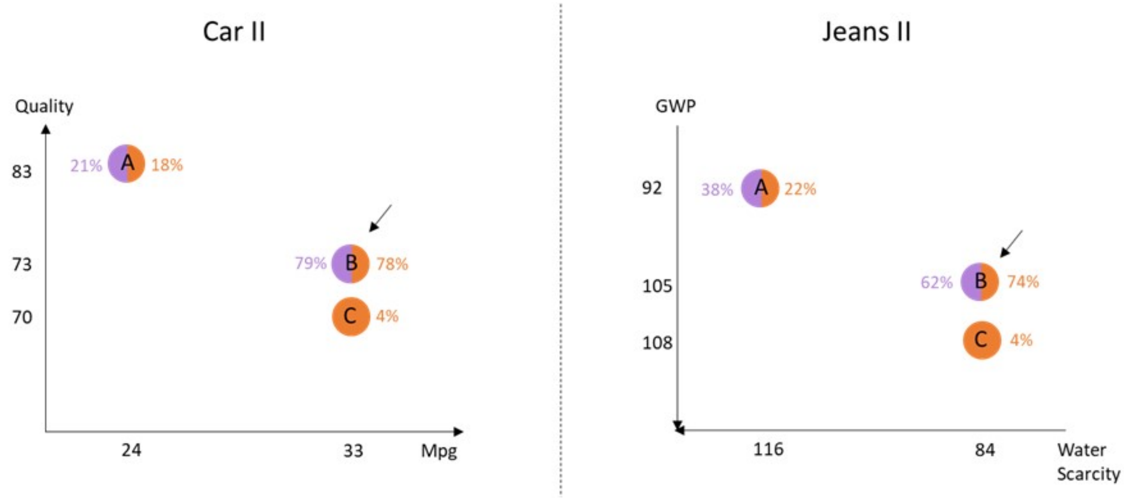


Figure A.4: Absolute choice proportions for the Car I and Jeans I experiments

A.4.2 Compromise Effect

The Figures A.5 - A.10 below illustrate the results for each pair of experiments for the attraction effect with their corresponding choice proportions.

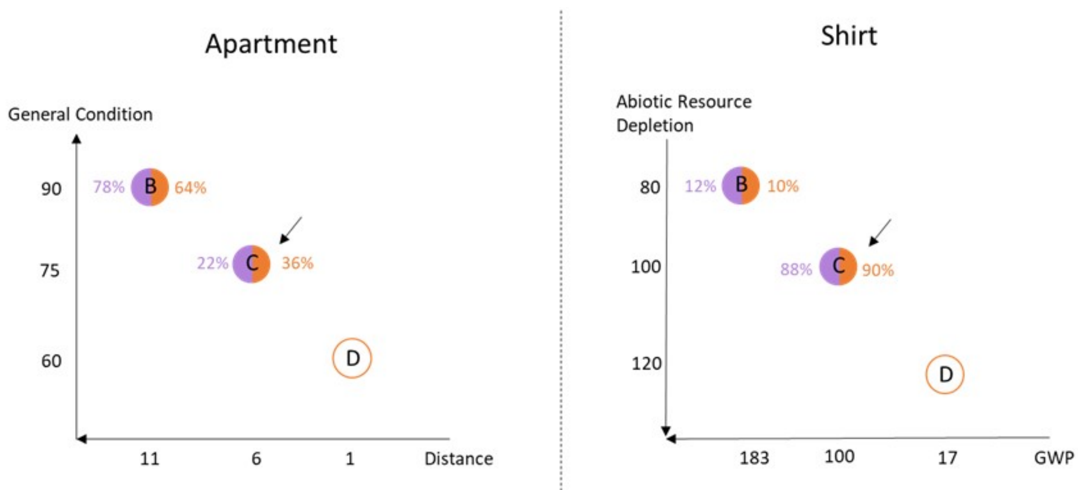


Figure A.5: Absolute choice proportions for the Apartment and Shirt experiments (alternative D was shown but not available to choose)

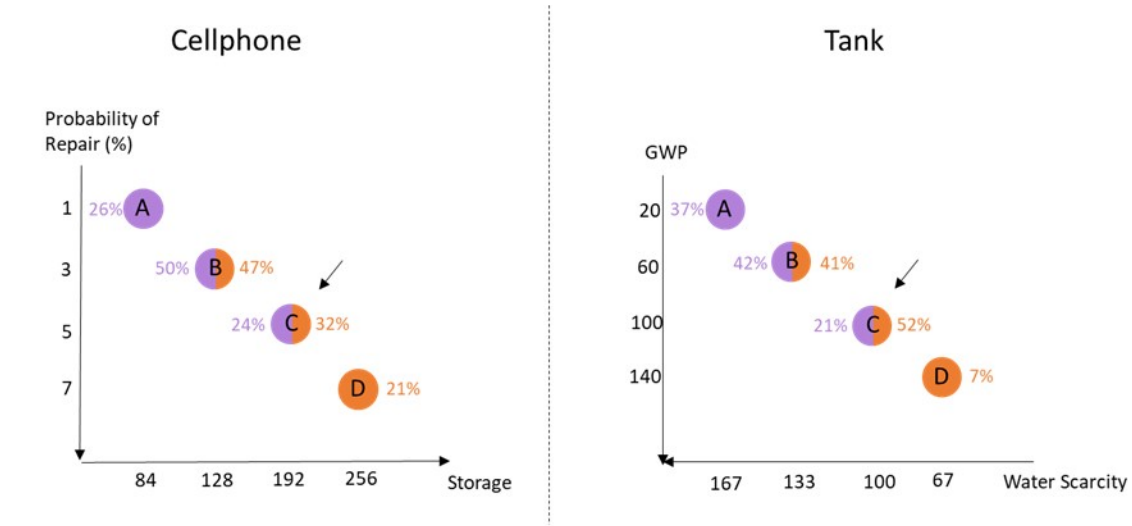


Figure A.6: Absolute choice proportions for the Cellphone and Tank experiments

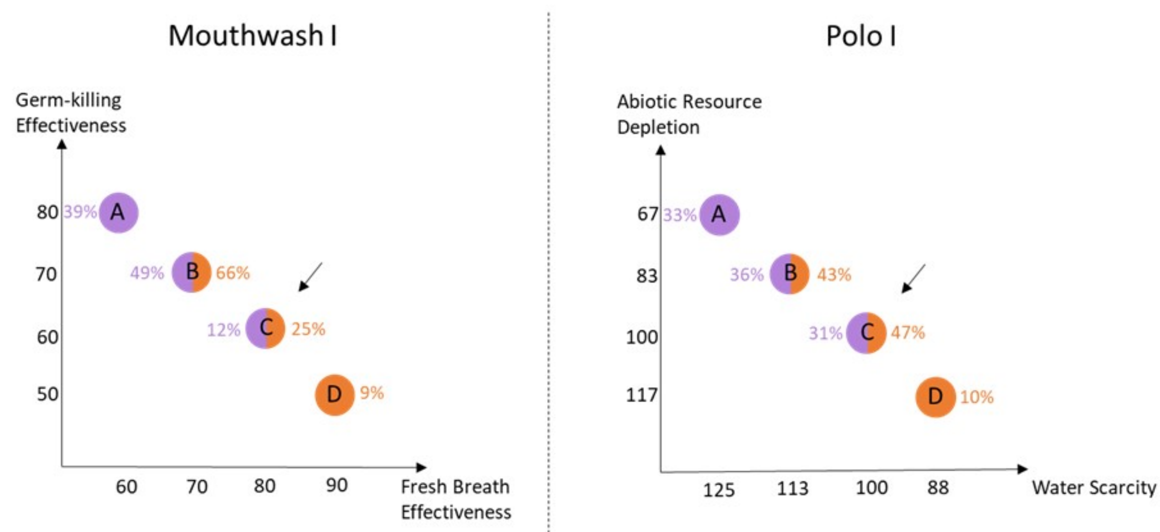


Figure A.7: Absolute choice proportions for the Mouthwash I and Polo I experiments

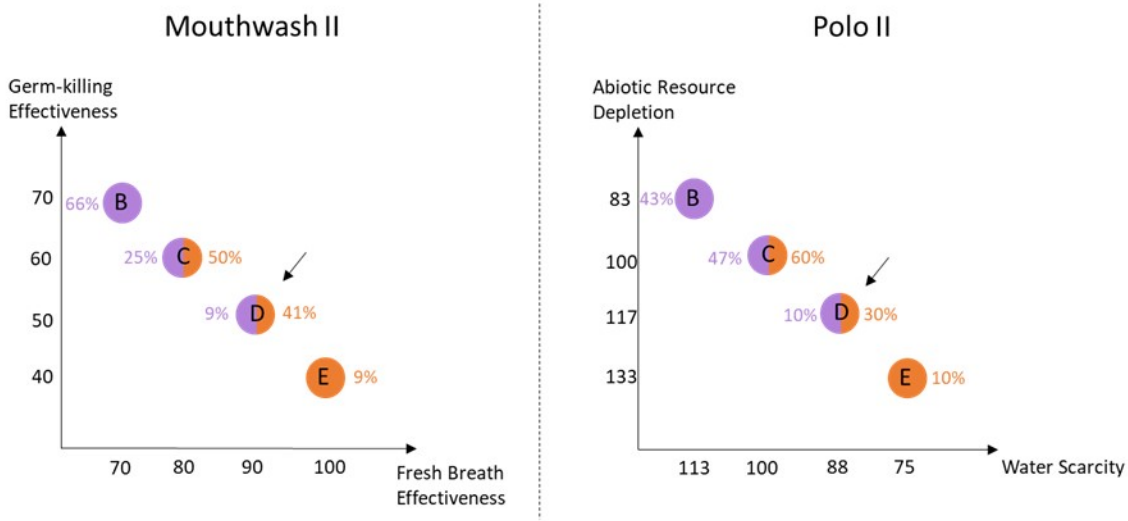


Figure A.8: Absolute choice proportions for the Mouthwash II and Polo II experiments

Figure A.9: Absolute choice proportions for the Batteries I and Chino I experiments (alternatives A and D were shown but not available to choose)

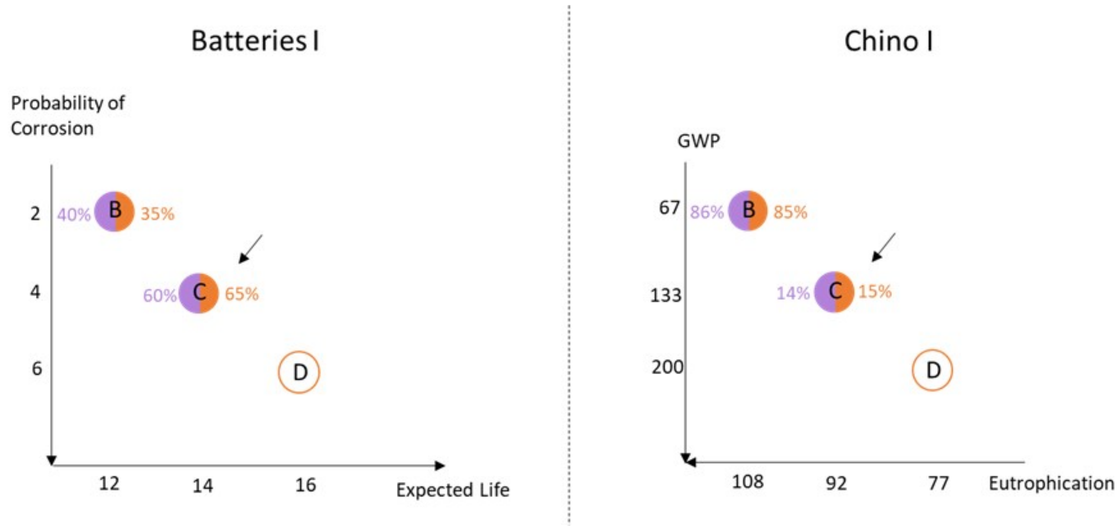


Figure A.10: Absolute choice proportions for the Batteries II and Chino II experiments (alternatives A and D were shown but not available to choose)

Note that all the choice proportions given in the figures above are the exact choice proportions for each treatment of each experiment. In the figures in the main text, and in the logistic regressions, we focus on the choice proportions of the alternatives that are present both in the core treatment and the treatment with decoy, in order to isolate the attraction and compromise effect. We refer to this as the conditional choice proportions. For an example of how this was done for an experiment involving the compromise effect, consider the experiment Mouthwash I. The alternatives that are common in both the core treatment and the treatment with decoy are B and C, our target option is C, and the decoy that is added in the decoy treatment is D. In the core treatment, 39% chose A, 49% chose B and 12% chose C. Then the conditional choice proportions become: 49 / (49+12) = 80% for B, and 12 / (49+12) = 20% for C. In the treatment with decoy D, 66% chose B, 25% chose C and 9% chose D. The conditional proportion for B is given by 66 / (66+25) = 73% and for C it is 25 / (66+25) = 27%. So, we observe a 7-percentage point increase in the conditional choice proportion of the target option C going from the core treatment to the treatment with decoy.

A.5 Statistical Analyses

To conduct statistical analyses, we performed logistic regression using R version 3.6.1. to determine whether the attraction or compromise effects are statistically significant. Our logistic regression equation is as follows:

$$Y_i = \beta_0 + \beta_1 C_i[+\beta_2(moderators)_i + \beta_3 C_i(moderators)_i] + \epsilon_i \tag{A.1}$$

The dependent variable $Y_i = +1$, if the target option is chosen and is 0 otherwise. The independent variable C_i indicates whether the treatment associated with response i involved a decoy or not. We used contrast coding (Judd et al. 2017) for C_i, i.e., $C_i = +1$ for the treatment with decoy and $C_i = -1$ for the core treatment.

By using contrast coding, the intercept becomes the mean of all observations included in the regression. We also used moderators such as age, gender, ethnicity, environmental knowledge, and environmental attitude, together with their interactions with the treatment contrast code. For environmental knowledge, we asked participants to agree/disagree with statements such as “I know more about eutrophication than the average person”. For environmental concern, we used the eco-centric concern questions from the environmental attitudes inventory by Milfont and Duckitt (2010). The full scales are included in Section A.6 below.

The output from the logistic regression analyses tests whether there is a statistically significant difference between the conditional choice proportions of the target option in each treatment, which is a two-sided hypothesis test. However, to say that we observe an attraction or compromise effect, the conditional choice proportion of the target option in the treatment with decoy should be higher than the conditional choice proportion in the core treatment, so we need a one-sided test. We convert the two-sided p-values from the regression output to one-sided p-values as follows. Let pr be the p-value from the regression, and p_{os} be the one-sided p-value that we seek. If the coefficient estimate is positive, we set $p_{os} = p_r/2$; if it is negative, we set $p_{os} = 1–p_r/2$. Whenever $p_{os} < 0.05$, we can say that we observe a statistically significant attraction/compromise effect in that experiment with significance level $p < 5\%$.

In addition, whenever we see a moderately significant (with $p_r < 0.10$) interaction effect between the treatment contrast code Ci and any of the moderators we used, we also report the moderator. If the coefficient of the interaction variable is positive, then the moderator had a positive effect on the magnitude of the attraction/compromise effect. The moderator variables created for controlling the effects of demographic information are as follows:

$$\textit{Gender1} = \begin{cases} 1, & \text{if Male} \\ 0, & \text{otherwise} \end{cases} \tag{A.2}$$

$$\textit{Gender2} = \begin{cases} 1, & \text{if Other} \\ 0, & \text{otherwise} \end{cases} \tag{A.3}$$

$$\textit{Ethnicity1} = \begin{cases} 1, & \text{if American Indian or Alaska Native} \\ 0, & \text{otherwise} \end{cases} \tag{A.4}$$

$$\textit{Ethnicity2} = \begin{cases} 1, & \text{if Asian} \\ 0, & \text{otherwise} \end{cases} \tag{A.5}$$

$$\textit{Ethnicity3} = \begin{cases} 1, & \text{if Black or African American} \\ 0, & \text{otherwise} \end{cases} \tag{A.6}$$

$$Ethnicity4 = \begin{cases} 1, & \text{if Other} \\ 0, & \text{otherwise} \end{cases} \tag{A.7}$$

$$Ethnicity5 = \begin{cases} 1, & \text{if (not White) and (all previous ethnicity variables are 0)} \\ 0, & \text{otherwise} \end{cases} \tag{A.8}$$

$$Knowledge = \text{average of the environmental knowledge questions on the scale from 1-7} \tag{A.9}$$

$$Attitude = \text{average of the environmental attitude questions on the scale from 1-7} \tag{A.10}$$

We standardized the continuous moderators (attitude, knowledge, and age) for each regression. Table A.13 summarizes our logistic regression results, and shows which moderators had a positive or negative effect on the attraction/compromise effect variable for each experiment. The abbreviations we used for the moderators in Table A.13 are as follows:

- age: age
- know: environmental knowledge score
- att: environmental attitude score
- G: gender (default is female, the most prevalent response); M: Male, O: Other
- E: ethnicity (default is white, the most prevalent response); AA: African American, AI: American Indian, A: Asian, O: Other, M: Multiple

As seen in Table A.13, we do not observe any clear pattern for the effect of the moderators.

Experiment	p_{os}	Positive Factors ($p_r < 0.10$)	Negative Factors ($p_r < 0.10$)	Sample Size
		Attraction effect experiments		
Soda I*	0.003			433
T-shirt I*	0		age, E:AA	432
Soda II	0.656			426
T-shirt II	0.18			420
Car I*	0	know	age, E:O	427
Jeans I	0.656			423
Car II	0.929	E:A, E:O, att		424
Jeans II*	0.001		E:A	424
		Compromise effect experiments		
Apartment*	0.004		E:AA	653
Shirt	0.403			653
Cellphone*	0.035		age, G:M,	337
Tank*	0.002			342
Mouthwash I*	0.001	att	E:O	335
Polo I*	0.007	age, know	G:M, att	347
Mouthwash II*	0.001	E:M	G:M	267
Polo II*	0.006			317
Batteries I*	0.022			433
Chino I	0.259	att		433
Batteries II*	0.046			439
Chino II	0.549	E:O		439

Table A.13: Summary of the one-sided logistic regression results for all experiments

Note that these results are obtained using the logistic regression specification. The experiments in which a statistically significant ($p_{os} < 0.05$) attraction or compromise effect were found are indicated with a "*" next to their name. Sample sizes differ within each experiment, as each participant only saw one treatment from each set of treatments associated with each experiment. Participants did see the equivalent treatments in the original experiment as in the transformed environmental version of the experiment. The reason why the sample sizes sometimes differ slightly between the replication of Simonson (1989)'s experiment and its environmental transformation is because we use conditional choice proportions in the regressions; the number of participants who selected one of the options from the core set could differ between the original treatment and the environmental transformation. To cast a wider net, moderators are reported if they are significant at $p_r < 0.10$, but even then no consistent patterns emerge.

So far, all results presented refer to the analyses of the full sample, where we combined the responses from all 4 subsamples: BLAB, MTurk, SustIS and SustIA. Below, we also present the conditional choice proportions of the target option in the core treatment and the treatment with decoy, and the magnitude of the difference (which indicates the magnitude of any attraction/compromise effect), for each subsample as well as for the combined sample in Tables A.14 - A.16. We combined the SustIS the SustIA subsamples as the SustIA alumni sample was too small to be analyzed separately.

Combined Data (n = 653)			
Experiment	core	with decoy	diff.
Soda I	0.49	0.79	0.3
T-shirt I	0.29	0.59	0.3
Soda II	0.51	0.5	-0.01
T-shirt II	0.71	0.74	0.03
Car I	0.21	0.37	0.15
Jeans I	0.38	0.37	0
Car II	0.79	0.82	0.03
Jeans II	0.62	0.77	0.15
Apartment	0.22	0.36	0.14
Shirt	0.88	0.9	0.01
Cellphone	0.32	0.4	0.16
Tank	0.34	0.56	0.35
Mouthwash I	0.2	0.28	0.15
Polo I	0.46	0.53	0.21
Mouthwash II	0.27	0.45	0.36
Polo II	0.17	0.33	0.24
Batteries I	0.6	0.65	0.05
Chino I	0.14	0.15	0.02
Batteries II	0.4	0.61	0.21
Chino II	0.86	0.86	0

Table A.14: Conditional choice proportions of the target alternative in core treatment and treatment with decoy, and their difference for all samples combined

	BLab (n = 230)			MTurk (n = 260)		
Experiment	core	with decoy	diff.	core	with decoy	diff.
Soda I	0.32	0.69	0.37	0.56	0.82	0.26
T-shirt I	0.22	0.52	0.3	0.36	0.67	0.31
Soda II	0.68	0.64	-0.05	0.44	0.35	-0.09
T-shirt II	0.78	0.83	0.06	0.64	0.6	-0.04
Car I	0.2	0.41	0.21	0.29	0.42	0.12
Jeans I	0.34	0.4	0.06	0.45	0.35	-0.1
Car II	0.8	0.88	0.07	0.71	0.73	0.03
Jeans II	0.66	0.79	0.13	0.55	0.74	0.19
Apartment	0.25	0.33	0.07	0.16	0.32	0.16
Shirt	0.92	0.96	0.04	0.84	0.81	-0.03
Cellphone	0.23	0.38	0.21	0.47	0.46	0.1
Tank	0.3	0.56	0.39	0.38	0.62	0.36
Mouthwash I	0.16	0.23	0.13	0.21	0.34	0.21
Polo I	0.5	0.42	0.1	0.52	0.65	0.29
Mouthwash II	0.19	0.41	0.35	0.34	0.46	0.31
Polo II	0.19	0.24	0.15	0.12	0.36	0.28
Batteries I	0.66	0.67	0	0.56	0.63	0.06
Chino I	0.09	0.17	0.08	0.21	0.17	-0.03
Batteries II	0.34	0.59	0.25	0.44	0.57	0.14
Chino II	0.91	0.91	0	0.79	0.79	0

Table A.15: Conditional choice proportions of the target alternative in core treatment and treatment with decoy, and their difference for BLab and MTurk subsamples

	SustIS + SustIA		
	(n = 163 (148 + 15))		
Experiment	core	with decoy	diff.
Soda I	0.62	0.87	0.25
T-shirt I	0.26	0.58	0.32
Soda II	0.38	0.53	0.15
T-shirt II	0.74	0.82	0.08
Car I	0.11	0.22	0.11
Jeans I	0.32	0.37	0.05
Car II	0.89	0.86	-0.03
Jeans II	0.68	0.78	0.1
Apartment	0.27	0.48	0.2
Shirt	0.9	0.94	0.04
Cellphone	0.21	0.33	0.19
Tank	0.3	0.47	0.28
Mouthwash I	0.24	0.25	0.12
Polo I	0.33	0.47	0.24
Mouthwash II	0.19	0.49	0.43
Polo II	0.24	0.42	0.29
Batteries I	0.58	0.66	0.08
Chino I	0.09	0.09	0
Batteries II	0.42	0.69	0.27
Chino II	0.91	0.91	0

Table A.16: Conditional choice proportions of the target alternative in core treatment and treatment with decoy, and their difference for SustIS and SustIA subsamples combined

As can be seen from Tables A.14 - A.16, our results within the subsamples are broadly consistent with the results from the full sample. The results from the full sample will typically yield higher significance levels due to the larger sample size, but the magnitudes of the effects are comparable within the subsamples.

A.6 Environmental Scales

The scales that we used for environmental knowledge (7-point scale, from strongly disagree (1) to strongly agree (7)) are as follows:

1. I know more about eutrophication than the average person.
2. I understand the environmental jargon associated with lifecycle assessments.
3. I am very knowledgeable about water pollution prevention techniques.
4. I am confident that I know how to select products and packages that reduce the amount of waste ending up in landfills.
5. I am knowledgeable about the consequences of fossil fuels depletion.

The scales that we used for environmental concern (attitude) (from Milfont and Duckitt (2010); 7-point scale, from strongly disagree (1) to strongly agree (7)) are as follows:

1. The idea that nature is valuable for its own sake is naive and wrong.
2. It makes me sad to see natural environments destroyed.
3. Nature is valuable for its own sake.
4. One of the worst things about overpopulation is that many natural areas are getting destroyed.
5. I do not believe protecting the environment is an important issue.

6. Despite our special abilities humans are still subject to the laws of nature.
7. It makes me sad to see forests cleared for agriculture.
8. It does NOT make me sad to see natural environments destroyed.
9. I do not believe nature is valuable for its own sake.
10. I don't get upset at the idea of forests being cleared for agriculture.

A.7 Limitations

This study was intended as a preliminary investigation of whether behavioral factors, such as context effects, manifest themselves when making choices involving environmental attributes. This study has several limitations, which point to natural opportunities for further research.

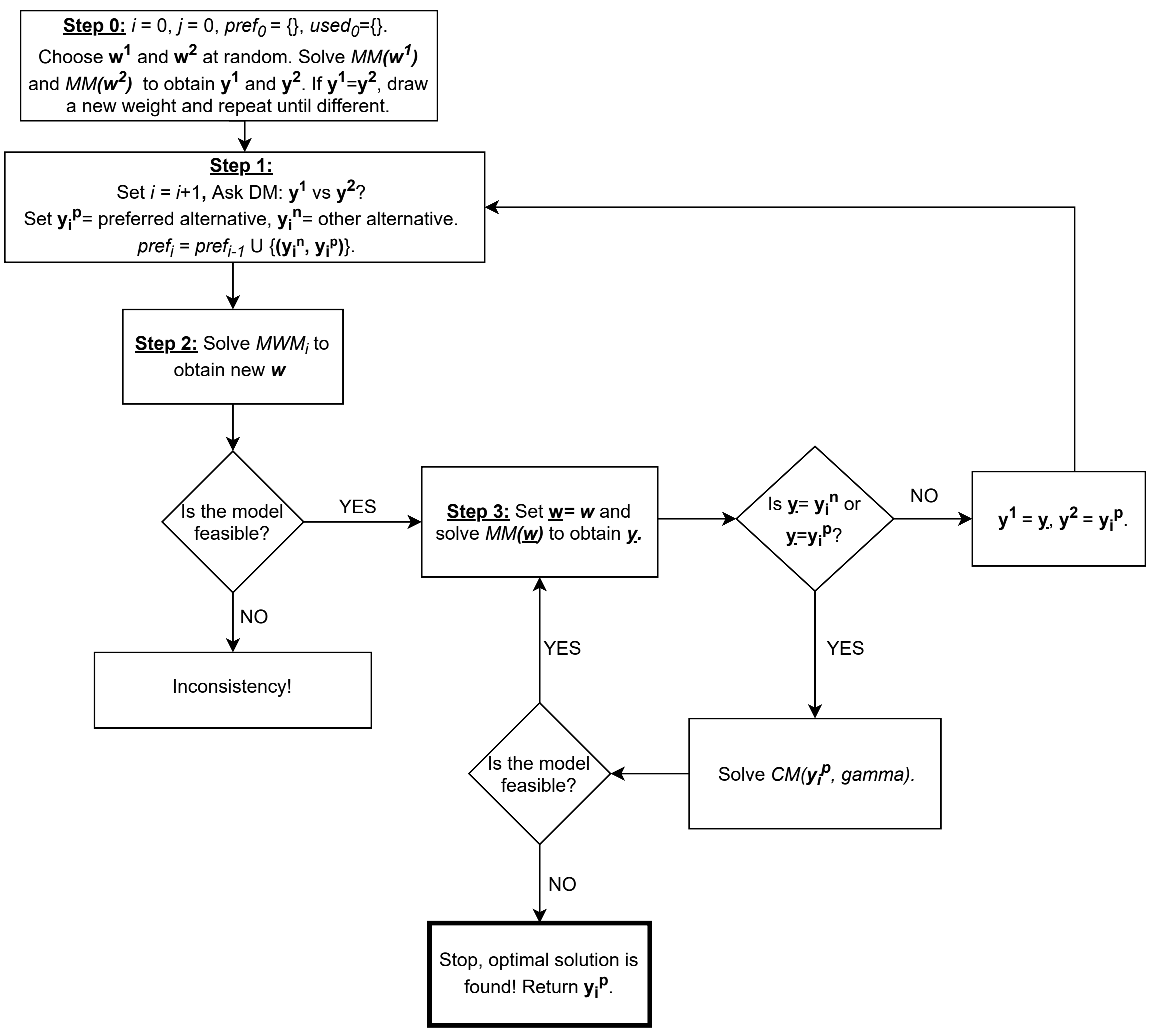
Our results are based on lab rather than field experiments. By including a population of environmentally knowledgeable participants we attempted to attain somewhat more external validity, but a field experiment involving decision-makers facing actual choices would be informative. We used Simonson (1989)'s classical consumer-oriented experiments as a benchmark against which to compare our results, but our environmental replications of those experiments asked participants to imagine themselves working as a designer in an apparel firm. It is conceivable that individuals would respond differently when primed as consumers vs. as decision-makers in an organization, though we did not find any evidence of that in our results. In designing the environmental experiments, we selected arbitrary pairs of environmental attributes (from the Higg MSI); the experiments were not designed specifically to examine systematic differences between the attributes, such as whether some attributes may be considered sacred or secular, or whether preferences related to attributes may be nonlinear. We included a few experiments with a social attribute (probability of child labor occurring in the supply chain) and one of the four earlier environmental attributes, but the framing of this social attribute was probably not sufficiently clear to participants and no consistent patterns emerged.

The final design of the experiments and construction of the Qualtrics survey tool took place during the initial stages of the Covid-19 pandemic, which limited the interaction between the research team. Two types of errors occurred, which are unfortunate, but which do not affect the results and findings presented here. First, due to a coding error in Qualtrics, the compromise effect experiment related to TV choice in Simonson (1989) and our environmental counterpart involving sweatshirts was not shown to the BLab and MTurk samples. We excluded those treatments from our analysis. Second, we had included a few treatments where information was presented visually, following Frederick et al. (2014)'s finding that attraction effects may disappear in that case. In those treatments the visual representations were flawed, leading to one alternative dominating the others; those experiments are also excluded here.

APPENDIX B

Appendix to Chapter 3

B.1 The Flowchart of the Proposed Interactive Preference-based Optimization Method



B.2 The Pseudo-code of the Proposed Interactive Preference-based Optimization Method

---

Algorithm 1 Proposed Interactive Preference-based Optimization Method

---

Set $i \leftarrow 0$; $\mathcal{P}_0^{\preceq} \leftarrow \emptyset; \delta \leftarrow 0.10^{-5}$

<u>**Step 0:**</u> Assign $\mathbf{w^1}$ and $\mathbf{w^2}$ randomly; solve MM($\mathbf{w^1}$) and MM($\mathbf{w^2}$) to obtain $\mathbf{y^1}$ and $\mathbf{y^2}$.

while $\mathbf{y^1} = \mathbf{y^2}$ **do**

  Choose $\mathbf{w'} \neq \mathbf{w^2}$; solve MM($\mathbf{w'}$) to obtain $\mathbf{y'}$; $\mathbf{y^2} \leftarrow \mathbf{y'}$

end while

<u>**Step 1:**</u> $i \leftarrow i{+}1$; Ask DM: $\mathbf{y^1} \preceq \mathbf{y^2}$?

if DM says YES **then**

  Set $\mathbf{y^p} \leftarrow \mathbf{y^2}$; $\mathbf{y^n} \leftarrow \mathbf{y^1}$

else

  Set $\mathbf{y^p} \leftarrow \mathbf{y^1}$; $\mathbf{y^n} \leftarrow \mathbf{y^2}$

end if

$\mathcal{P}_i^{\preceq} \leftarrow \mathcal{P}_{i-1}^{\preceq} \bigcup \{(\mathbf{y^n}, \mathbf{y^p})\}$

<u>**Step 2:**</u> Solve MWM($\mathcal{P}_i^{\preceq}$) to obtain $\mathbf{w}$

if MWM($\mathcal{P}_i^{\preceq}$) is feasible **then**

  go to **Step 3**

else

  Stop, DM is inconsistent

end if

Step 3: Set $\underline{\mathbf{w}} \leftarrow \mathbf{w}$; solve MM($\underline{\mathbf{w}}$) to obtain $\underline{\mathbf{y}}$

if $\underline{\mathbf{y}} \neq \mathbf{y^p}$ & $\underline{\mathbf{y}} \neq \mathbf{y^n}$ **then**

Set $\mathbf{y^1} \leftarrow \underline{\mathbf{y}}$, $\mathbf{y^2} \leftarrow \mathbf{y^p}$ and go to **Step 1**

else

Solve CM($\mathcal{P}_i^{\preceq}, \mathbf{y^p}, \delta$) to obtain $\mathbf{w}$

if CM($\mathcal{P}_i^{\preceq}, \mathbf{y^p}, \delta$) is feasible **then**

Go to **Step 3**

else

Stop, optimal solution is found

end if

end if

return $\mathbf{y^p}$

APPENDIX C

Appendix to Chapter 4

C.1 The Experimental Flow for Treatment 1

Below, we provide the complete experimental flow for treatment 1. Treatments 2,3, and 4 will have either different ordering of the methods and/or different ordering of the attributes.

INTRODUCTION AND CONSENT:

You are invited to voluntarily participate in a research project conducted by Charles Corbett, Professor of Operations Management and Sustainability at UCLA on Prolific. Your participation in this study is voluntary.

Why is this study being done? We are conducting this study to understand how individuals make decisions in a business context using two different decision making methods.

What will happen if I take part in this research study? If you volunteer to participate in this study, the researcher will ask you to do the following:

(a) Provide your consent to participate.

(b) Use two different decision making methods to solve a business problem.

(c) Answer questions regarding your experiences with the above methods and demographics.

How long will I be in the research study? The survey will take about 20 minutes, including the consent process.

Are there any potential risks or discomforts that I can expect from this study?

We do not foresee any risks or discomfort.

Are there any potential benefits if I participate? You will contribute to the body of knowledge regarding decision-making and sustainability.

Will I receive any payment if I participate in this study? You will receive $5.00 for completing this study. You will receive the payment through Prolific.

Will information about me and my participation be kept confidential? Only the investigators will have access to your responses during the study. Personally identifiable information that Prolific has about you will not be shared with the investigators. Your responses to the survey questions will be entirely anonymous, and cannot be traced back to you. These anonymous data may be shared with other investigators.

What are my rights if I take part in this study? You may withdraw your consent at any time and discontinue participation without penalty. You can choose whether or not you want to be in this study. If you volunteer to be in this study, you may leave the study at any time without consequences of any kind.

Who can answer questions I might have about this study? If you have any questions, comments or concerns about the research, please contact Prof. Charles Corbett at charles.corbett@anderson.ucla.edu. If you have questions about your rights as a research subject, or you have concerns or suggestions and you want to talk to someone other than the researchers, you may contact the UCLA OHRPP by phone: (310) 206-2040; by email: participants@research.ucla.edu or by mail: Box 951406, Los Angeles, CA 90095-1406.

By continuing with the survey you indicate that you have read and understood the information provided above, and that you willingly agree to participate in this research study. Thank you for your willingness to participate.

CONTEXT AND GENERAL INSTRUCTIONS:

The apparel industry produces more than 30 million tons of goods annually and it accounts for more than 1700 million tons of carbon dioxide (CO_2) emissions, which contribute to climate change. One single t-shirt accounts for between 5-10 kg CO2. Cotton is one of the main raw materials in the textile industry. It takes approximately 2700 liters of water to make a single cotton t-shirt, which is enough drinking water for a person for 900 days. It is crucial to understand the environmental impacts of the apparel industry and make responsible decisions. These decisions are often hard, because they involve trade-offs between multiple economic and environmental considerations.

Several methods exist to provide guidance in making those hard decisions. In this experiment, you will be taking the role of an analyst who works in the apparel industry and who makes decisions based on metrics such as cost, global warming potential and water use. These metrics are also called your decision criteria. Cost will be expressed in thousands of dollars. Global warming potential and water use are expressed in scores, where a higher score means higher impacts; in other words, a higher score is worse for the environment. These scores are generated based on a state-of-the-art database of environmental impacts of the apparel industry. In all scenarios that you will see, the actual global warming potential and water use are substantial.

In this experiment, there is a demand for 10 different product types that needs to be satisfied. There are 15 different raw materials that can be supplied from 20 different suppliers around the world.

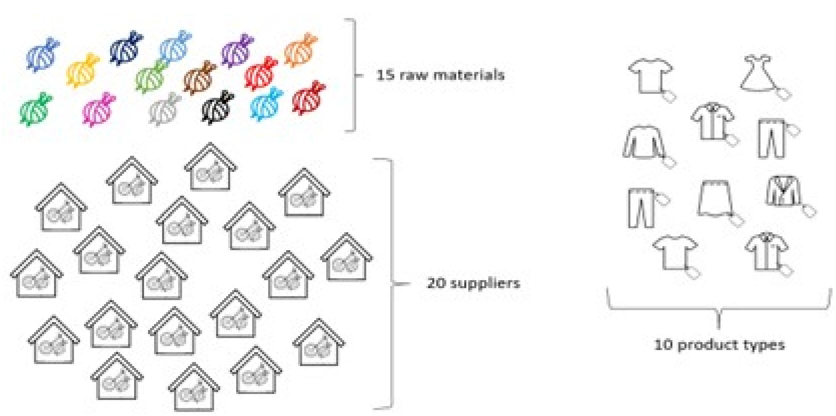


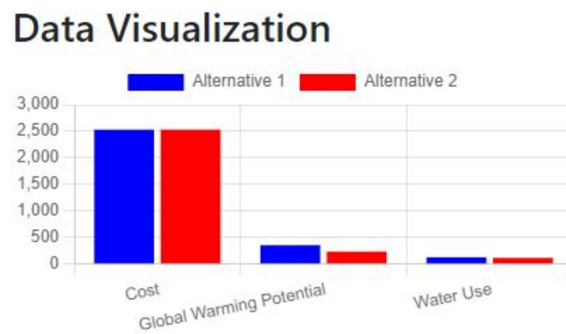
Figure C.1: Parameters of the Sustainable Sourcing Problem

The total cost, global warming potential and water use will vary depending on which raw materials are sourced from which suppliers. Your task will be to satisfy demand with a global sourcing strategy that yields the best combination of cost, global warming potential and water use, according to your preferences. The number of potential solutions is very large and it is complicated to make these decisions without any guidance.

Figure C.2: Illustration of Complicated Decision Making in the Sustainble Sourcing Problem

To illustrate the type of choice you will be making, please see the below comparison of two possible sourcing strategies. One of the alternatives is clearly better than the other. Please indicate the alternative which performs better.

Title	Alternative 1	Alternative 2
Cost	2531.26	2531.26
Global Warming Potential	356.21	236.21
Water Use	128.62	116.86



Choose an option below:

- Alternative 1
- Alternative 2

Continue

Figure C.3: First Attention Check Question

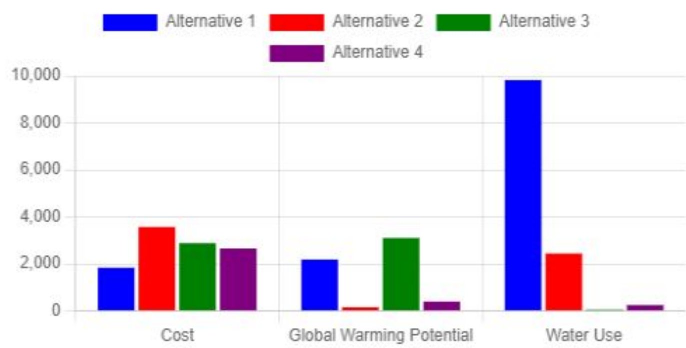
You will use two different decision-making tools to help with your decisions. One of them is called Direct Rating while the other is called Interactive Algorithm. In the Direct Rating method, you will be asked to rate the importance of each decision criterion, namely cost, global warming potential and water use. In the Interactive Algorithm, you will be presented with a series of pairwise comparison questions and asked to choose which alternative you prefer by evaluating their performance on cost, global warming potential and water use. After completing both methods, you will be presented with a survey to evaluate your experience with the two methods, followed by a few more general questions.

METHOD: DIRECT RATING

This decision-making method is called Direct Rating. Below we will show a representative set of examples of feasible solutions and their respective cost, global warming potential and water use. Remember that for all decision criteria, lower values are better. You are asked to rate each decision criterion, namely cost, global warming potential and water use, between 0-100 to indicate the importance of each criterion according to your personal preferences. A higher rating means that criterion is more important to you. The ratings do not need to add up to 100. Please rate each of the decision criteria below. You can change your ratings until you click continue.

Title	Alternative 1	Alternative 2	Alternative 3	Alternative 4
Cost	1840.41	3575.68	2884.62	2661.18
Global Warming Potential	2190.5	162.05	3107.3	401.01
Water Use	9825.61	2445.4	59.13	251.35

Data Visualization



Enter a value between 0-100 as the rating for Cost

Enter a value between 0-100 as the rating for Global Warming Potential

Enter a value between 0-100 as the rating for Water Use

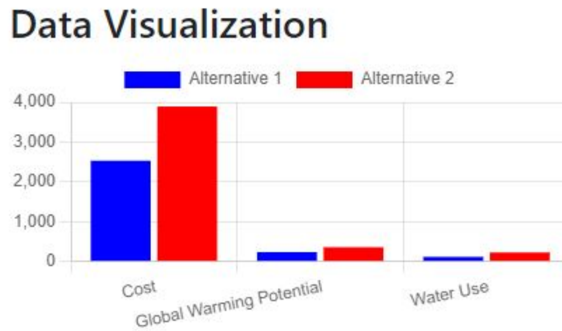
Continue

Figure C.4: Direct Rating

Thank you for submitting your ratings. You have entered: (input ratings) . The Direct Rating method can now use these ratings to find the optimal solution to the sourcing problem. We will show that optimal solution later.

Now please indicate which of the following two alternatives is better:

Title	Alternative 1	Alternative 2
Cost	2531.26	3892.87
Global Warming Potential	236.21	358.74
Water Use	116.86	227.16



Choose an option below:

- Alternative 1
- Alternative 2

Continue

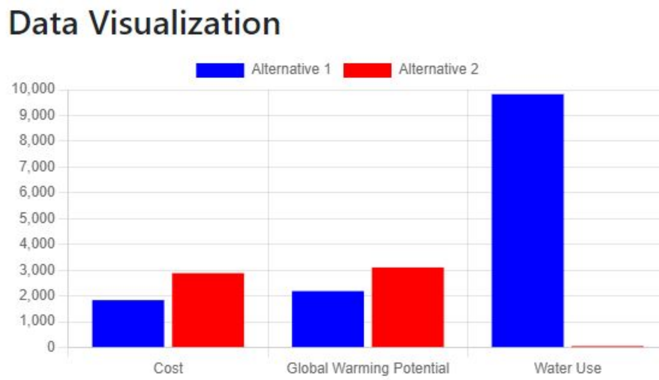
Figure C.5: Second Attention Check Question

METHOD: INTERACTIVE ALGORITHM:

This decision-making method is called Interactive Algorithm. In each iteration, we will show you two alternatives and we will ask you which one you prefer. The alternatives will be displayed in a table and as a bar chart. The Interactive Algorithm uses your answers in each round to decide which alternatives to show next. After you go to the next round, you cannot go back and change your response from the previous round. In each iteration, the Interactive Algorithm is trying to find a feasible alternative that might be better than the one you preferred in the previous round. The algorithm will stop once there is no such challenger left to be found.

There are two alternatives to choose from: Alternative 1 and Alternative 2. Their resulting cost, global warming potential and water use are indicated in the below table and illustrated in the below bar chart. Which one do you prefer: Alternative 1 or Alternative 2? Please indicate your choice below and submit.

Title	Alternative 1	Alternative 2
Cost	1840.41	2884.62
Global Warming Potential	2190.5	3107.3
Water Use	9825.61	59.13



Choose an option below:

- Alternative 1
- Alternative 2

Next

Figure C.6: Interactive Algorithm

INTERACTIVE ALGORITHM

The INTERACTIVE ALGORITHM is trying to find a challenger based on your preferences. Please wait!

Figure C.7: Wait Page to Find Challenger

Results for INTERACTIVE ALGORITHM

The algorithm could not find any challengers based on your preferences. The last alternative you chose is the optimal one:

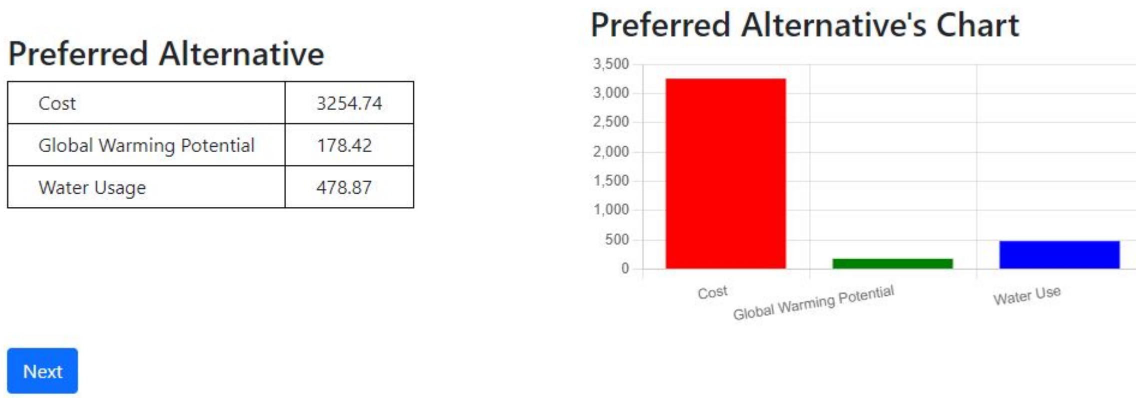


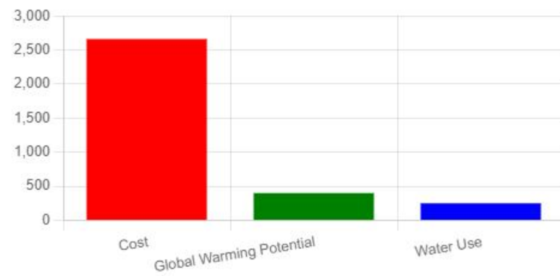
Figure C.8: Optimal Solution for Interactive Algorithm

COMPARING THE SOLUTIONS REACHED BY TWO METHODS:

Thank you for using both Direct Rating and Interactive Algorithm. For the Direct Rating, we took the ratings you provided for the three attributes (cost, global warming potential and water use), and determined the corresponding optimal solution to the sustainable sourcing problem. Below you can find the optimal solution reached with each method.

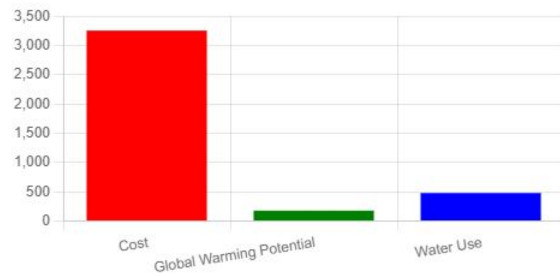
OPTIMAL SOLUTION WITH DIRECT RATING

Cost	2661.18
Global Warming Potential	401.01
Water Usage	251.35



OPTIMAL SOLUTION WITH INTERACTIVE ALGORITHM

Cost	3254.74
Global Warming Potential	178.42
Water Usage	478.87



Please indicate your preferences:

- ◯ I strongly prefer the solution with Interactive Algorithm
- ◯ I prefer the solution with Interactive Algorithm
- ◯ I somewhat prefer the solution with Interactive Algorithm
- ◯ I am indifferent between the two solutions
- ◯ I somewhat prefer the solution with Direct Rating
- ◯ I prefer the solution with Direct Rating
- ◯ I strongly prefer the solution with Direct Rating

Next

Figure C.9: Comparing Optimal Solutions

Please answer the following questions to tell us more about your experience with these methods.

(The survey questions to be used with 7-point Likert Scale: Strongly Agree - Agree - Somewhat Agree - Neither Agree nor Disagree - Somewhat Disagree - Disagree - Strongly Disagree)

Overall, I am satisfied with the Direct Rating.

Overall, I am satisfied with the Interactive Algorithm.

Direct Rating was easy to use.

Interactive Algorithm was easy to use.

It was easy to provide Direct Ratings.

The comparison questions in the Interactive Algorithm were easy to answer.

Using Direct Rating improved my understanding of my own values.

Using Interactive Algorithm improved my understanding of my own values.

Using Direct Rating improved my understanding of key trade-offs between alternatives.

Using Interactive Algorithm improved my understanding of key trade-offs between alternatives.

Using Direct Rating could be helpful in better communicating my decision-making results.

Using Interactive Algorithm could be helpful in better communicating my decision-making results.

I trust the outcome with Direct Rating.

I trust the outcome with Interactive Algorithm.

I am satisfied with the final outcome of the Direct Rating.

I am satisfied with the final outcome of the Interactive Algorithm.

I would like to try applying Direct Rating to other decision problems.

I would like to try applying Interactive Algorithm to other decision problems.

Thank you for reading thoroughly. Please select strongly disagree for this question.

Demographic questions:

What is your age?

What is your gender? (The choices are: Male, Female, Other, Prefer not to say)

Which country are you located in? (The choices are: US, UK)

Please select your highest degree or level of school you have completed from the drop down menu. (The choices are: Some High School, High School, Assoc Degree, Bachelors, Some Grad School, Masters, Other Prof. Degree, PhD, Other, Prefer not to say)

Information about your household income is very helpful when analyzing survey responses. Please select the range closest to your annual household income before taxes from the below drop-down menu. (The choices are: Less than $10,000, Between $10,000 and $25,000, Between $25,000 and $50,000, Between $50,000 and $75,000, Between $75,000 and $100,000, Between $100,000 and $150,000, More than $150,000, Prefer not to answer)

(The survey questions to be used with 7-point Likert Scale for environmental attitude:)

I think of myself as an environmentally friendly individual.

I would be willing to accept cuts in my standard of living to protect the environment.

Bibliography

Aksoy Y, Butler TW, Minor ED (1996) Comparative studies in interactive multiple objective mathematical programming. *European Journal of Operational Research* 89(2):408–422, ISSN 0377-2217.

Alavi B, Tavana M, Mina H (2021) A dynamic decision support system for sustainable supplier selection in circular economy. *Sustainable Production and Consumption* 27:905–920, ISSN 2352-5509, URL http://dx.doi.org/https://doi.org/10.1016/j.spc.2021.02.015.

Alptekinoğlu A, Örsdemir A (2022) Is adopting mass customization a path to environmentally sustainable fashion? *Manufacturing & Service Operations Management* 24(6):2982–3000, URL http://dx.doi.org/10.1287/msom.2022.1088.

Aubert AH, Esculier F, Lienert J (2020) Recommendations for online elicitation of swing weights from citizens in environmental decision making. *Operations Research Perspectives* 7:100156, ISSN 2214-7160, URL http://dx.doi.org/https://doi.org/10.1016/j.orp.2020.100156.

Bateman IJ, Munro A, Poe GL (2008) Decoy effects in choice experiments and contingent valuation: Asymmetric dominance. *Land Economics* 84(1):115–127, ISSN 0023-7639, URL http://dx.doi.org/10.3368/le.84.1.115.

Baumann H (2000) Introduction and organisation of lca activities in industry. *The International Journal of Life Cycle Assessment* 5(6):363–368.

Beaudrie C, Corbett CJ, Lewandowski TA, Malloy T, Zhou X (2021) Evaluating the application of decision analysis methods in simulated alternatives assessment case studies: Potential benefits and challenges of using mcda. *Integrated Environmental Assessment and Management* 17(1):27–41, URL http://dx.doi.org/https://doi.org/10.1002/ieam.4316.

Beemsterboer S, Baumann H, Wallbaum H (2020) Ways to get work done: a review and systematisation of simplification practices in the lca literature. *The International Journal of Life Cycle Assessment* 25(11):2154–2168.

Benayoun R, De Montgolfier J, Tergny J, Laritchev O (1971) Linear Programming with Multiple

Objective Functions: Step Method (STEM). *Mathematical Programming* 1(1):366–375.

Bertsimas D, O'Hair A (2013) Learning Preferences Under Noise and Loss Aversion: An Optimization Approach. *Operations Research* 61(5):1190–1199.

Bevilacqua M, Ciarapica FE, Mazzuto G, Paciarotti C (2014) Environmental Analysis of a Cotton Yarn Supply Chain. *Journal of Cleaner Production* 82:154 – 165, ISSN 0959-6526.

Bottomley PA, Doyle JR, Green RH (2000) Testing the reliability of weight elicitation methods: Direct rating versus point allocation. *Journal of Marketing Research* 37(4):508–513, URL `http://dx.doi.org/10.1509/jmkr.37.4.508.18794`.

Büyüközkan G, Çifçi G (2011) A Novel Fuzzy Multi-criteria Decision Framework for Sustainable Supplier Selection with Incomplete Information. *Computers in Industry* 62(2):164 – 174, ISSN 0166-3615.

Carlsson F, Gravert C, Johansson-Stenman O, Kurz V (2021) The use of green nudges as an environmental policy instrument. *Review of Environmental Economics and Policy* 15(2):216–237, URL `http://dx.doi.org/10.1086/715524`.

Carlsson F, Mørkbak MR, Olsen SB (2012) The first time is the hardest: A test of ordering effects in choice experiments. *Journal of Choice Modelling* 5(2):19–37, ISSN 1755-5345, URL `http://dx.doi.org/https://doi.org/10.1016/S1755-5345(13)70051-4`.

Ceballos B, Lamata MT, Pelta DA (2016) A comparative analysis of multi-criteria decision-making methods. *Progress in Artificial Intelligence* 5(4):315–322.

Chen DL, Schonger M, Wickens C (2016) otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance* 9:88–97, ISSN 2214-6350, URL `http://dx.doi.org/https://doi.org/10.1016/j.jbef.2015.12.001`.

Choi S, Kim J, Lee E, Lee J (2022) Probability weighting and cognitive ability. *Management Science* 68(7):5201–5215, URL `http://dx.doi.org/10.1287/mnsc.2021.4146`.

Cinelli M, Kadziński M, Miebs G, Gonzalez M, Słowiński R (2022) Recommending multiple criteria decision analysis methods with a new taxonomy-based decision support system. *European Journal of Operational Research* 302(2):633–651, ISSN 0377-2217, URL `http://dx.doi.org/https://doi.org/10.1016/j.ejor.2022.01.011`.

Cinelli M, Miebs G (2022) Find the most relevant weighting methods for your decision-making problem with our WEighting Methods Selection Software (WEMSS). https://mcda.cs.put.poznan.pl/wemss/index.php, online; Last Accessed: 08-May-2024.

Croson R, Treich N (2014) Behavioral environmental economics: Promises and challenges. *Environmental and Resource Economics* 58(3):335–351.

Dai T, Tang C (2022) Frontiers in service science: Integrating esg measures and supply chain management: Research opportunities in the postpandemic era. *Service Science* 14(1):1–12, URL http://dx.doi.org/10.1287/serv.2021.0295.

Davis AM, Huang R, Thomas DJ (2022) Retailer inventory sharing in two-tier supply chains: An experimental investigation. *Management Science* 68(12):8773–8790, URL http://dx.doi.org/10.1287/mnsc.2022.4323.

Deb K, Sinha A, Korhonen PJ, Wallenius J (2010) An interactive evolutionary multiobjective optimization method based on progressively approximated value functions. *IEEE Transactions on Evolutionary Computation* 14(5):723–739, URL http://dx.doi.org/10.1109/TEVC.2010.2064323.

Drobner C, Goerg SJ (2024) Motivated belief updating and rationalization of information. *Management Science* null, URL http://dx.doi.org/10.1287/mnsc.2023.02537.

Drolet A, Luce MF, Jiang L, Rossi BC, Hastie R (2020) The Preference for Moderation Scale. *Journal of Consumer Research* 47(6):831–854, ISSN 0093-5301, URL http://dx.doi.org/10.1093/jcr/ucaa042.

Drolet A, Luce MF, Simonson I (2008) When Does Choice Reveal Preference? Moderators of Heuristic versus Goal-Based Choice. *Journal of Consumer Research* 36(1):137–147, ISSN 0093-5301, URL http://dx.doi.org/10.1086/596305.

Dyer JS (1973) A Time-Sharing Computer Program for the Solution of the Multiple Criteria Problem. *Management Science* 19(12):1379–1383.

Dyer JS, Fishburn PC, Steuer RE, Wallenius J, Zionts S (1992) Multiple Criteria Decision Making, Multiattribute Utility Theory: The Next Ten Years. *Management Science* 38(5):645–654.

Dyer JS, Sarin RK (1979) Measurable Multiattribute Value Functions. *Operations Research* 27(4):810–822.

Erhun F, Kraft T, Wijnsma S (2021) Sustainable triple-a supply chains. *Production and Operations Management* 30(3):644–655, URL http://dx.doi.org/https://doi.org/10.1111/poms.13306.

Frederick S, Lee L, Baskin E (2014) The limits of attraction. *Journal of Marketing Research* 51(4):487–507, URL http://dx.doi.org/10.1509/jmr.12.0061.

Galindro BM, Welling S, Bey N, Olsen SI, Soares SR, Ryding SO (2020) Making use of life cycle assessment and environmental product declarations: A survey with practitioners. *Journal of Industrial Ecology* 24(5):965–975, URL http://dx.doi.org/https://doi.org/10.1111/jiec.13007.

Geoffrion AM, Dyer JS, Feinberg A (1972) An Interactive Approach for Multi-Criterion Optimization, with an Application to the Operation of an Academic Department. *Management Science* 19(4-part-1):357–368.

Global Fashion Agenda, Boston Consulting Group (2017) Pulse of the Fashion Industry. https://globalfashionagenda.com/wp-content/uploads/2017/05/Pulse-of-the-Fashion-Industry_2017.pdf, online; Last Accessed: 12-August-2020.

Govindan K, Khodaverdi R, Jafarian A (2013) A Fuzzy Multi Criteria Approach for Measuring Sustainability Performance of a Supplier Based on Triple Bottom Line Approach. *Journal of Cleaner Production* 47:345 – 354, ISSN 0959-6526.

Grubert E (2017) Implicit prioritization in life cycle assessment: text mining and detecting metapatterns in the literature. *The International Journal of Life Cycle Assessment* 22(2):148–158.

Guérin-Schneider L, Tsanga-Tabi M, Roux P, Catel L, Biard Y (2018) How to better include environmental assessment in public decision-making: Lessons from the use of an lca-calculator for wastewater systems. *Journal of Cleaner Production* 187:1057–1068, ISSN 0959-6526, URL http://dx.doi.org/https://doi.org/10.1016/j.jclepro.2018.03.168.

Guinée J, Heijungs R (2017) *Introduction to Life Cycle Assessment*, 15–41 (Cham: Springer International Publishing), ISBN 978-3-319-29791-0.

Hämäläinen RP (2015) Behavioural issues in environmental modelling – the missing perspective. *Environmental Modelling Software* 73:244–253, ISSN 1364-8152, URL `http://dx.doi.org/https://doi.org/10.1016/j.envsoft.2015.08.019`.

Hämäläinen RP, Alaja S (2008) The threat of weighting biases in environmental decision analysis. *Ecological Economics* 68(1):556–569, ISSN 0921-8009, URL `http://dx.doi.org/https://doi.org/10.1016/j.ecolecon.2008.05.025`.

Hicks AL, Gilbertson LM, Yamani JS, Theis TL, Zimmerman JB (2015) Life cycle payback estimates of nanosilver enabled textiles under different silver loading, release, and laundering scenarios informed by literature review. *Environmental Science & Technology* 49(13):7529–7542, URL `http://dx.doi.org/10.1021/acs.est.5b01176`.

Hobbs B, Meier P (1994) Multicriteria methods for resource planning: an experimental comparison. *IEEE Transactions on Power Systems* 9(4):1811–1817, URL `http://dx.doi.org/10.1109/59.331435`.

Hodgett RE (2016) Comparison of multi-criteria decision-making methods for equipment selection. *The International Journal of Advanced Manufacturing Technology* 85(5):1145–1157.

Hofstetter P, Lippiatt BC, Bare JC, Rushing AS (2002) User Preferences for Life-Cycle Decision Support Tools: Evaluation of a Survey of BEES Users. `https://tsapps.nist.gov/publication/get`$_p df.cfm?pub_i d = 860066, online; LastAccessed : 29 - April - 2024$.

Hofstetter P, Mettier TM (2003) What users want and may need. *Journal of Industrial Ecology* 7(2):79–101, URL `http://dx.doi.org/https://doi.org/10.1162/108819803322564361`.

Huber J, Payne JW, Puto C (1982) Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *Journal of Consumer Research* 9(1):90–98, ISSN 0093-5301, URL `http://dx.doi.org/10.1086/208899`.

Ifcher J, Zarghamee H (2023) Does decision making for others close the gender gap in competition? *Management Science* null, URL `http://dx.doi.org/10.1287/mnsc.2023.4861`.

Ishizaka A, Siraj S (2018) Are multi-criteria decision-making tools useful? an experimental comparative study of three methods. *European Journal of Operational Research* 264(2):462–471, ISSN 0377-2217, URL `http://dx.doi.org/https://doi.org/10.1016/j.ejor.2017.05.041`.

Judd C, McClelland G, Ryan C (2017) *Data Analysis: A Model Comparison Approach To Regression, ANOVA, and Beyond Third Edition (3rd ed.).* (Routledge).

Jónasson JO, Ramdas K, Sungu A (2022) Social impact operations at the global base of the pyramid. *Production and Operations Management* 31(12):4364–4378, URL `http://dx.doi.org/https://doi.org/10.1111/poms.13857`.

Karimi L, Yacuel L, Degraft-Johnson J, Ashby J, Green M, Renner M, Bergman A, Norwood R, Hickenbottom KL (2022) Water-energy tradeoffs in data centers: A case study in hot-arid climates. *Resources, Conservation and Recycling* 181:106–194, ISSN 0921-3449, URL `http://dx.doi.org/https://doi.org/10.1016/j.resconrec.2022.106194`.

Kesternich M, Reif C, Rübbelke D (2017) Recent trends in behavioral environmental economics. *Environmental and Resource Economics* 67(3):403–411.

Köksalan M, Wallenius J (2012) *Multiple Criteria Decision Making: Foundations and Some Approaches*, chapter Chapter 9, 171–183. URL `http://dx.doi.org/10.1287/educ.1120.0097`.

Köksalan MM, Sagala PNS (1995) Interactive approaches for discrete alternative multiple criteria decision making with monotone utility functions. *Management Science* 41(7):1158–1171, URL `http://dx.doi.org/10.1287/mnsc.41.7.1158`.

Korhonen P, Silvennoinen K, Wallenius J, Öörni A (2012) Can a Linear Value Function Explain Choices? An Experimental Study. *European Journal of Operational Research* 219(2):360 – 367, ISSN 0377-2217.

Korhonen P, Soleimani-damaneh M, Wallenius J (2016) Dual cone approach to convex-cone dominance in multiple criteria decision making. *European Journal of Operational Research* 249(3):1139–1143, ISSN 0377-2217, URL `http://dx.doi.org/https://doi.org/10.1016/j.ejor.2015.09.043`.

Korhonen P, Wallenius J, Zionts S (1984) Solving the Discrete Multiple Criteria Problem using Convex Cones. *Management Science* 30(11):1336–1345.

Laudenbach C, Ungeheuer M, Weber M (2023) How to alleviate correlation neglect in investment decisions. *Management Science* 69(6):3400–3414, URL http://dx.doi.org/10.1287/mnsc.2022.4535.

Linkov I, Seager TP (2011) Coupling multi-criteria decision analysis, life-cycle assessment, and risk assessment for emerging threats. *Environmental Science & Technology* 45(12):5068–5074, URL http://dx.doi.org/10.1021/es100959q.

Lokman B, Köksalan M, Korhonen PJ, Wallenius J (2016) An interactive algorithm to find the most preferred solution of multi-objective integer programs. *Annals of Operations Research* 245(1):67–95, URL http://dx.doi.org/10.1007/s10479-014-1545-2.

Lokman B, Köksalan M, Korhonen PJ, Wallenius J (2018) An interactive approximation algorithm for multi-objective integer programs. *Computers & Operations Research* 96:80–90, ISSN 0305-0548, URL http://dx.doi.org/https://doi.org/10.1016/j.cor.2018.04.005.

Luo Y, Song K, Ding X, Wu X (2021) Environmental sustainability of textiles and apparel: A review of evaluation methods. *Environmental Impact Assessment Review* 86:106497, ISSN 0195-9255, URL http://dx.doi.org/https://doi.org/10.1016/j.eiar.2020.106497.

Mackin PD, Roy A, Wallenius J (2011) An interactive weight space reduction procedure for nonlinear multiple objective mathematical programming. *Mathematical Programming* 127(2):425–444, URL http://dx.doi.org/10.1007/s10107-009-0293-6.

Manzardo A, Ren J, Piantella A, Mazzi A, Fedele A, Scipioni A (2014) Integration of Water Footprint Accounting and Costs for Optimal Chemical Pulp Supply Mix in Paper Industry. *Journal of Cleaner Production* 72.

McKone TE, Nazaroff WW, Berck P, Auffhammer M, Lipman T, Torn MS, Masanet E, Lobscheid A, Santero N, Mishra U, Barrett A, Bomberg M, Fingerman K, Scown C, Strogen B, Horvath A (2011) Grand challenges for life-cycle assessment of biofuels. *Environmental Science & Technology* 45(5):1751–1756, URL http://dx.doi.org/10.1021/es103579c.

Memari A, Dargi A, Akbari Jokar MR, Ahmad R, Abdul Rahim AR (2019) Sustainable Supplier Selection: A Multi-criteria Intuitionistic Fuzzy TOPSIS Method. *Journal of Manufacturing Systems* 50:9 – 24, ISSN 0278-6125.

Mendoza Beltran A, Prado V, Font Vivanco D, Henriksson PJG, Guinée JB, Heijungs R (2018) Quantified uncertainties in comparative life cycle assessment: What can be concluded? *Environmental Science & Technology* 52(4):2152–2161, URL http://dx.doi.org/10.1021/acs.est.7b06365.

Mettier T, Scholz R (2008) Measuring preferences on environmental damages in lcia. part 2: choice and allocation questions in panel methods. *The International Journal of Life Cycle Assessment* 13(6):468–476.

Mettier T, Scholz R, Tietje O (2006) Measuring preferences on environmental damages in lcia. part 1: Cognitive limits in panel surveys. *The International Journal of Life Cycle Assessment* 11(6):394–402.

Mierlo KV, Rohmer S, Gerdessen JC (2017) A Model for Composing Meat Replacers: Reducing the Environmental Impact of our Food Consumption Pattern while Retaining its Nutritional Value. *Journal of Cleaner Production* 165:930 – 950, ISSN 0959-6526.

Milfont TL, Duckitt J (2010) The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudes. *Journal of Environmental Psychology* 30(1):80–94, ISSN 0272-4944, URL http://dx.doi.org/https://doi.org/10.1016/j.jenvp.2009.09.001.

Moazzem S, Crossin E, Daver F, Wang L (2022) Environmental impact of apparel supply chain and textile products. *Environment, Development and Sustainability* 24(8):1–19.

Morana R, Seuring S (2011) A Three Level Framework for Closed-Loop Supply Chain Management-Linking Society, Chain and Actor Level. *Sustainability* 3(4):678–691, ISSN 2071-1050.

Mousseau V, Figueira J, Naux JP (2001) Using assignment examples to infer weights for electre tri method: Some experimental results. *European Journal of Operational Research* 130(2):263–275, ISSN 0377-2217, URL http://dx.doi.org/https://doi.org/10.1016/S0377-2217(00)00041-2.

Muthu SS (2020) *Introduction to sustainability and the textile supply chain and its environmental impact*, 13 (Woodhead Publishing).

Muthu SS, Li Y, Hu J, Mok P (2012) Quantification of Environmental Impact and Ecological Sustainability for Textile Fibres. *Ecological Indicators* 13(1):66 – 74, ISSN 1470-160X.

Muthulingam S, Corbett CJ, Benartzi S, Oppenheim B (2013) Energy efficiency in small and medium-sized manufacturing firms: Order effects and the adoption of process improvement recommendations. *Manufacturing & Service Operations Management* 15(4):596–615, URL http://dx.doi.org/10.1287/msom.2013.0439.

Nielsen KS, Brick C, Hofmann W, Joanes T, Lange F, Gwozdz W (2022) The motivation–impact gap in pro-environmental clothing consumption. *Nature Sustainability* 5(8):665–668.

Oppenheimer KR (1978) A Proxy Approach to Multi-Attribute Decision Making. *Management Science* 24(6):675–689.

Oppewal H, Huybers T, Crouch GI (2015) Tourist destination and experience choice: A choice experimental analysis of decision sequence effects. *Tourism Management* 48:467–476, ISSN 0261-5177, URL http://dx.doi.org/https://doi.org/10.1016/j.tourman.2014.12.016.

Osbaldiston R, Schott JP (2012) Environmental sustainability and behavioral science: Meta-analysis of proenvironmental behavior experiments. *Environment and Behavior* 44(2):257–299, URL http://dx.doi.org/10.1177/0013916511402673.

Öztürk BA, Özçelik F (2014) Sustainable Supplier Selection with A Fuzzy Multi-Criteria Decision Making Method Based on Triple Bottom Line. *Business and Economics Research Journal* 5(3):129–147.

Pfister S, Bayer P, Koehler A, Hellweg S (2011) Environmental impacts of water use in global crop production: hotspots and trade-offs with land use. *Environmental Science & Technology* 45(13):5761–5768.

Phelps SP, Köksalan M (2003) An Interactive Evolutionary Metaheuristic for Multiobjective Combinatorial Optimization. *Management Science* 49(12):1726–1738.

Pizzolo M, Laurent A, Sala S, Weidema B, Verones F, Koffler C (2017) Normalisation and weighting in life cycle assessment: Quo vadis? *The International Journal of Life Cycle Assessment* 22(6):853–866, ISSN 0948-3349, URL http://dx.doi.org/10.1007/s11367-016-1199-1.

Pryshlakivsky J, Searcy C (2021) Life cycle assessment as a decision-making tool: Practitioner and managerial considerations. *Journal of Cleaner Production* 309:127344, ISSN 0959-6526, URL http://dx.doi.org/https://doi.org/10.1016/j.jclepro.2021.127344.

Radhakrishnan S (2015) *The Sustainable Apparel Coalition and the Higg Index*, 23–57 (Singapore: Springer Singapore), ISBN 978-981-287-164-0.

Rajagopal D, Vanderghem C, MacLean HL (2017) Life cycle assessment for economists. *Annual Review of Resource Economics* 9:361–381, ISSN 1941-1359, URL `http://dx.doi.org/https://doi.org/10.1146/annurev-resource-100815-095513`.

Reap J, Roman F, Duncan S, Bras B (2008) A survey of unresolved problems in life cycle assessment. *The International Journal of Life Cycle Assessment* 13(5):374–388.

Rex E, Baumann H (2008) Implications of an interpretive understanding of lca practice. *Business Strategy and the Environment* 17(7):420–430, URL `http://dx.doi.org/https://doi.org/10.1002/bse.633`.

Rowley H, Peters G, Lundie S, Moore S (2012) Aggregating Sustainability Indicators: Beyond the Weighted Sum. *Journal of Environmental Management* 111:24–33.

Roy A, Wallenius J (1992) Nonlinear Multiple Objective Optimization: An Algorithm and Some Theory. *Mathematical Programming* 55:235–249.

SAC (2020) Sustainable Apparel Coalition - Higg MSI. `https://msi.higg.org/sac-materials/1/textiles`, online; Last Accessed: 05-June-2020.

Sandin G, Peters GM, Svanström M (2013) Moving Down the Cause-effect Chain of Water and Land Use Impacts: An LCA Case Study of Textile Fibres. *Resources, Conservation and Recycling* 73:104 – 113, ISSN 0921-3449.

Shen B (2014) Sustainable Fashion Supply Chain: Lessons from H&M. *Sustainability* 6(9):6236–6249, ISSN 2071-1050.

Shin WS, Ravindran A (1991) Interactive Multiple Objective Optimization: Survey I-Continuous Case. *Computers & Operations Research* 18(1):97 – 114, ISSN 0305-0548.

Simonson I (1989) Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research* 16(2):158–174.

Sinha A, Korhonen P, Wallenius J, Deb K (2014) An interactive evolutionary multi-objective optimization algorithm with a limited number of decision maker calls.

European Journal of Operational Research 233(3):674–688, ISSN 0377-2217, URL `http://dx.doi.org/https://doi.org/10.1016/j.ejor.2013.08.046`.

Stewart TJ (1993) Use of Piecewise Linear Value Functions in Interactive Multicriteria Decision Support: A Monte Carlo Study. *Management Science* 39(11):1369–1381.

Stillwell WG, Barron F, Edwards W (1983) Evaluating credit applications: A validation of multiattribute utility weight elicitation techniques. *Organizational Behavior and Human Performance* 32(1):87–108, ISSN 0030-5073, URL `http://dx.doi.org/https://doi.org/10.1016/0030-5073(83)90141-1`.

Sunar N, Swaminathan JM (2022) Socially relevant and inclusive operations management. *Production and Operations Management* 31(12):4379–4392, URL `http://dx.doi.org/https://doi.org/10.1111/poms.13873`.

Testa F, Nucci B, Tessitore S, Iraldo F, Daddi T (2016) Perceptions on lca implementation: evidence from a survey on adopters and nonadopters in italy. *The International Journal of Life Cycle Assessment* 21(10):1501–1513.

Thies C, Kieckhäfer K, Spengler TS, Sodhi MS (2019) Operations Research for Sustainability Assessment of Products: A Review. *European Journal of Operational Research* 274(1):1 – 21, ISSN 0377-2217.

Tickner JA, Schifano JN, Blake A, Rudisill C, Mulvihill MJ (2015) Advancing safer alternatives through functional substitution. *Environmental Science & Technology* 49(2):742–749.

Toffano F, Garraffa M, Lin Y, Prestwich S, Simonis H, Wilson N (2022) A multi-objective supplier selection framework based on user-preferences. *Annals of Operations Research* 308(1):609–640, URL `http://dx.doi.org/10.1007/s10479-021-04251-5`.

Toubia O, Hauser JR, Simester DI (2004) Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis. *Journal of Marketing Research* 41(1):116–131.

Toubia O, Simester DI, Hauser JR, Dahan E (2003) Fast Polyhedral Adaptive Conjoint Estimation. *Marketing Science* 22(3):273–303.

Tseng S, Hung S (2014) A Strategic Decision-Making Model Considering the Social Costs of Carbon Dioxide Emissions for Sustainable Supply Chain Management. *Journal of Environmental Management* 133:315 – 322, ISSN 0301-4797.

van der Velden N, Patel M, Vogtländer J (2014) LCA Benchmarking Study on Textiles Made of Cotton, Polyester, Nylon, Acryl, or Elastane. *International Journal of Life Cycle Assessment* 19:331 – 356.

Velez MA, Moros L (2021) Have behavioral sciences delivered on their promise to influence environmental policy and conservation practice? *Current Opinion in Behavioral Sciences* 42:132–138, URL `http://dx.doi.org/https://doi.org/10.1016/j.cobeha.2021.06.008`.

Walker MJ, Katok E, Shachat J (2023) Trust and trustworthiness in procurement contracts with retainage. *Management Science* 69(6):3492–3515, URL `http://dx.doi.org/10.1287/mnsc.2022.4516`.

Wallenius J (1975) Comparative Evaluation of Some Interactive Approaches to Multicriterion Optimization. *Management Science* 21(12):1387–1396.

Wallenius J, Dyer JS, Fishburn PC, Steuer RE, Zionts S, Deb K (2008) Multiple Criteria Decision Making, Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead. *Management Science* 54(7):1336–1349.

Walser T, Demou E, Lang DJ, Hellweg S (2011) Prospective environmental life cycle assessment of nanosilver t-shirts. *Environmental Science & Technology* 45(10):4570–4578, URL `http://dx.doi.org/10.1021/es2001248`.

Wang L, Shen B (2017) A Product Line Analysis for Eco-Designed Fashion Products: Evidence from an Outdoor Sportswear Brand. *Sustainability* 9(7):1136, ISSN 2071-1050.

Wang M, Yang J (1998) A multi-criterion experimental comparison of three multi-attribute weight measurement methods. *Journal of Multi-Criteria Decision Analysis* 7(6):340–350, URL `http://dx.doi.org/https://doi.org/10.1002/(SICI)1099-1360(199811)7:6<340::AID-MCDA206>3.0.`

Wu JJ, Mazzuchi TA, Sarkani S (2023) Comparison of multi-criteria decision-making methods for online controlled experiments in a launch decision-making framework. *Information and Software Technology* 155:107115, ISSN 0950-5849, URL `http://dx.doi.org/https://doi.org/10.1016/j.infsof.2022.107115`.

WWF (2020) World Wild Life Sustainabile Agriculture - Cotton. `https://www.worldwildlife.org/industries/cotton`, online; Last Accessed: 01-May-2020.

Zanakis SH, Solomon A, Wishart N, Dublish S (1998) Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research* 107(3):507–529, ISSN 0377-2217, URL `http://dx.doi.org/https://doi.org/10.1016/S0377-2217(97)00147-1`.

Zionts S (1981) A Multiple Criteria Method for Choosing Among Discrete Alternatives. *European Journal of Operational Research* 7(2):143 – 147, ISSN 0377-2217.

Zionts S, Wallenius J (1976) An Interactive Programming Method for Solving the Multiple Criteria Problem. *Management Science* 22(6):652–663.

Zionts S, Wallenius J (1980) Identifying Efficient Vectors: Some Theory and Computational Results. *Operations Research* 28(3-part-ii):785–793.

Zionts S, Wallenius J (1983) An Interactive Multiple Objective Linear Programming Method for a Class of Underlying Nonlinear Utility Functions. *Management Science* 29(5):519–529.