

Product Lines, Product Design and Limited Attention*

Carsten Dahremöller[†] and Markus Fels[‡]

7th April, 2012

Abstract

We analyze how firms design their product lines when facing customers with limited attention. We assume that consumers simplify complex decision problems by neglecting several of the relevant aspects. Whether and to what extent a customer pays attention to an attribute of a product depends on the importance of the attribute as well as its dispersion in the set of alternatives. A firm may thus influence its customers' attention through the range of products it makes available. We show that a firm can increase its profit by introducing additional goods that have the sole function of manipulating consumer attention. Hence, even with a homogenous consumer population our model can explain product differentiation. We derive several results on how a firm can effectively manipulate its customers' attention.

JEL Classification: D40, D80, L10, L12, L15

Keywords: Limited Attention, Saliency, Focus, Product Design, Product Lines

*In preparing this paper we have greatly benefited from comments made by Paul Heidhues, Daniel Krämer, and Matthias Kräkel. We also thank the seminar audience in Bonn. All remaining errors are our own responsibility. This work was supported by the Bonn Graduate School of Economics.

[†]Bonn Graduate School of Economics, University of Bonn, Kaiserstrasse 1, 53113 Bonn, Germany; E-mail address: dahremoeller@uni-bonn.de

[‡]Bonn Graduate School of Economics, University of Bonn, Kaiserstrasse 1, 53113 Bonn, Germany; E-mail address: s3mafels@uni-bonn.de

1 Introduction

We propose a model of limited attention and analyze its implications for product design. As most consumer products have several characteristics, any purchase decision is a complex problem of trading off advantages in some dimensions against disadvantages in others. We posit that the way in which a decision-maker pays attention to different aspects of the problem reflects a need to simplify such complex decisions. The resulting attention allocation reflects both the decision-maker's valuation of the different aspects of a product and the extent to which the available options (products and outside option) differ in a dimension. We investigate how a firm optimally designs its product line when facing customers whose attention is determined in such a way. We start with the optimal design of a single product that is offered by a monopolist. The optimal design will not only reflect the extent to which customers value a particular characteristic, but will also reflect to which extent customers consider a particular attribute (the attention allocation) and how a particular design changes this attention allocation. We then investigate how a firm can profit from expanding its product line. We show that there is an incentive to offer multiple products to a set of customers with homogeneous preferences. This incentive to offer differentiated products is based on the fact that the attention allocation is determined by the choice set. Finally, we discuss the optimal design of a product line.

We show that limited attention tends to decrease the complexity of the offered products. Products are simpler in the sense that they are equipped with fewer features and hence are easier to grasp for consumers. The most central proposition of the paper is that firms offer two kinds of products. There is a primary good that is intended to be sold to consumers. In addition, firms produce what we call bait goods. These are not intended to be sold but have the sole purpose of manipulating consumer attention in a way that is favorable to the firm.¹ We find that the optimal attention manipulation weighs incentives to redirect attention to more profitable attributes against incentives to maximize attention paid to each attribute of the product. In contrast to much general perception, it turns out that the firm tends to suffer from consumers having limited attention.

¹The assumption that a bait good is only used to manipulate consumer attention is useful to simplify the analysis. However, we will later present some extensions in which the bait good can also be profitably sold to some consumers. One example would be luxury goods that are sold to a small population of rich consumers and at the same time serve as a bait good to the larger population of less wealthy consumers.

This is due to the fact that limited attention tends to decrease the willingness-to-pay of consumers, which in turn lowers the price that firms can charge for a given product. While customers may not fully consider the price either, we show that a firm is not able to exploit this inattention.

In our model, each product has a set of attributes. For example, a car can be equipped with features like a sunroof, electric windows, or with certain levels of horsepower or safety. Apart from these rather salient characteristics there are many others that are less noticeable. Examples would be the average durability of the gear box, the seating comfort or the design of the cigarette lighter. Hence, such complex products like a car have a plethora of different attributes that all add to the overall quality of the product. Although information is freely available for lots of these attributes, consumers usually focus on a few key variables. Such behavior of systematic neglect of important information can arguably be classified as limited attention.

We join an emerging strand of the economic literature that incorporates such limited attention into economic models. Gabaix (2011) develops a model of limited attention in which the attention allocation is the solution to an optimal attention allocation problem. Our paper draws heavily on his work, as the attention allocation we assume can be derived from an attention allocation problem with cognitive costs similar to the one proposed by Gabaix (2011). Our modeling of attention allocation differs from his approach as we employ some modified assumptions we deem appropriate to reflect the costs of complexity. A more detailed comparison is deferred to later.² In addition, the focus of our paper are the implications of limited attention on market outcomes like product design and price. Kőszegi and Szeidl (2011) propose a model of focus-weighted utility which exhibits some parallels to our model. They show that a firm has an incentive to concentrate the advantages of its products in few dimensions while spreading the disadvantages across as many dimensions as possible. While such an incentive is also present in our model (yet for different reasons, as we will explain in the next section), we derive more detailed results regarding the optimal product design. Spiegel (2006a) and Spiegel (2006b) study the optimal industry behavior if consumers act according to the $S(1)$ sampling routine developed by Osborne and Rubinstein (1998). Beyond the attention heuristic employed we deem the most important difference to be the welfare implications of limited attention. In Spiegel (2006a) and Spiegel

²The intuition for the attention allocation is given in the main text. A detailed derivation of it can be found in the Appendix.

(2006b) customers can be exploited by firms which obfuscate their products. Similarly, Rubinstein (1993) describes a firm's incentive to use complex price schemes to extract additional profits from boundedly rational customers. In contrast, we find limited attention to primarily hurt the firm while benefiting customers, despite a firm's ability to manipulate attention. Eliaz and Spiegler (2011) propose a model that also features products whose sole function is to attract attention. Consumers only consider the products of a subset of the firms in the market. Therefore, a firm uses attention grabbers if it wants consumers to consider its products. In contrast we investigate a firm's potential to attract or distract attention from product characteristics (including the price), thereby manipulating the desirability of a purchase. Bordalo, Gennaioli, and Shleifer (2010) and Bordalo (2011) develop a framework of limited attention to account for choice set effects. Their idea of limited attention is inspired by psychological findings concerning the perception of alternatives. While these papers focus on the explanation of several experimental results, we seek to develop predictions concerning the impact of limited attention - which in our framework refers to cognitive restrictions, not errors in perception - on market outcomes.

There is a large literature on choice set effects (Simonson 1989; Huber, Payne, and Puto 1982) and their impact on behavior in various settings (Herne 1997; McFadden 1999; Benartzi and Thaler 2002). Several explanations for compromise effects³ have been proposed - ranging from extreme-ness aversion (Simonson and Tversky 1992) to information inference from choice sets (Wernerfelt 1995; Kamenica 2008). Kamenica (2008) shows that information inference produces an incentive to offer premium loss leaders. Though not explicitly relating his model to the compromise effect Vikander (2010) proposes a model of prestige goods and describes a firm's incentives to advertise premium products to an audience which is not able to afford the purchase. As an example, Vikander (2010) presents some anecdotes of advertisement campaigns for premium goods. For instance, Audi advertised its \$118,000 R8 in the half-time of Super Bowl XLII. This advertisement spot cost Audi six million dollars. What is puzzling about this story is that only a minority of Super Bowl viewers are able to afford such a car. So presumably if Audi wanted to advertise this car to people who are able to buy it, there would be other far less costly marketing channels to reach the particular target audience. We will show that in our model as well there is an incentive to offer

³The compromise effect depicts that expanding the choice set by a product which is more extreme in one attribute than any of the previously available options makes products which are mediocre in that attribute look more favorably.

premium products which are not intended for sale. Yet in our framework this incentive is based on a firm's ability to manipulate its customers' attention, and not on a firm's attempt to signal product value (Kamenica 2008) or increase its products' prestige value (Vikander 2010).

The remainder of this paper will proceed as follows. In section 2 we introduce and analyze the underlying attention process. Section 3 is dedicated to the derivation of the optimal product design if only one product can be supplied. This will be extended in section 4 where firms can introduce an additional product that is designed to manipulate consumer attention. Section 5 analyzes the effects of a firm offering several of these manipulating products. We continue by discussing some possible extensions in section 6. Section 7 concludes.

2 The Role of Attention

This section introduces the model that we use to depict the customers' way of deciding between alternatives in their choice set. A main feature is that there is a difference between *experienced utility* and *decision utility*. Experienced utility measures the satisfaction or welfare customers derive from a choice. It depends on the attributes an alternative features and the value a customer assigns to these attributes. In contrast, decision utility depicts the way in which customers choose between alternatives. Decision utility thus not only depends on the welfare a customer derives from an alternative, but also on the way choices are made, and thus encompasses choice procedures, perceptions of alternatives (at the moment of choice), salience of attributes and alternatives, and the like. The distinction is supposed to depict the contrast between utility as a measure of welfare and utility as a tool to model choice (behavior).⁴ In our model, the salience of attributes results in a difference between experienced and decision utility. The reason behind this discrepancy is the limited cognitive ability of humans to decide optimally on the plethora of information that is available to them. To be able to focus solely on the limited ability to process information, we do not model any costs of obtaining information or costs of searching. Instead we will assume that all product information is readily available, but that consumers have problems of converting product information into an overall assessment of desirability.

⁴For a distinction of the concepts of experienced utility and decision utility, see Kahneman and Tversky (1984) and Kahneman (2000).

We assume experienced utility to take a simple structure. An alternative's attributes are perfect substitutes. Formally, let the experienced utility of an alternative $a \in A$ be denoted by

$$u(a) = \sum_{i=1}^n v_i x_i^a.$$

The term $v_i \in \mathbb{R}$ measures the (marginal) value of attribute i to the decision maker (henceforth DM). The variable x_i^a measures the extent to which alternative a features attribute i . We will assume that $v_i \neq v_{i'}, \forall i \neq i'$.

The consumer does not base her choice on experienced utility, but on decision utility. Formally, the decision utility of an alternative is a function of the experienced utility and the salience of each attribute:

$$\tilde{u}(a) = \sum_{i=1}^n m_i v_i x_i^a. \tag{1}$$

The term $m_i \in [0, 1]$ is the attention parameter associated with attribute i . If $m_i = 0$, attribute i is completely neglected. In this case, any differences between alternatives in attribute i will be irrelevant for the decision. We normalize the attention such that $m_i \leq 1 \forall i$.⁵ Since attention is normalized such that $m_i \leq 1$, an attention allocation of $m_i = 1$ means that the DM pays full attention to attribute i . Note here that we would be back in the rational model if $m_i = m_j > 0 \forall i, j$.

We now want to motivate and analyze the attention process of the customer that determines the attention parameters m_i . First, note that we do not model a problem of optimal attention allocation. While such a problem may, to some extent, underly the way in which attention is distributed, we assume it to be given at the moment of choice. Instead, we assume a particular rule how the salience parameters m_i are determined and give empirical and analytical reasons why this rule is sensible. We then seek to investigate how a firm designs its products when faced with customers that apply this way of allocating attention.⁶

⁵Note that this normalization is without loss of generality since a decision utility with some attention vector \mathbf{m} yields the same choice behavior as a decision utility with an attention vector $\alpha \cdot \mathbf{m}$ for $\alpha > 0$.

⁶The attention allocation that we assume is motivated by several decision characteristics that are suggested by experimental evidence. In addition, the proposed attention allocation can in fact be derived as the solution to an attention allocation problem (see the Appendix). Hence, the particular attention allocation is optimal for

We do not intend to model perceptual mistakes. The DM is able to perfectly determine differences between alternatives in each dimension. The limitations in the cognitive process arise when the DM needs to integrate the information about differences between alternatives in multiple dimensions with his own evaluation concerning the importance of each dimension in order to reach an evaluation of each alternative. This task usually includes making judgments as to how an advantage in one dimension trades off against a disadvantage in another dimension. As the number of relevant dimensions increases, so does the cognitive effort of evaluating all resulting trade offs. We seek to model one way in which a decision-maker may deal with such a complex decision problem, namely by simplification through neglect and prioritizing of aspects of the decision problem.⁷

We want the attention allocation to reflect a need to simplify a complex problem in order to be able to solve it. We thus want the attention allocation to follow some basic rules:

- (a) The level of attention that a DM allocates to an attribute rises with its importance to the DM (v_i).
- (b) The more the alternatives that the DM faces differ in a given attribute, i.e. the larger the contrast within one dimension, the higher is the attention that the DM allocates to that attribute.
- (c) The cost of considering another attribute rises with the number of attributes that are already considered.

We will now give a short overview of the assumptions and their impact on the attention process.

Assumption (a) is present in most models of limited attention. The DM allocates more attention to those dimensions he deems more important. When making a decision he tends to focus on those dimensions from which he derives most value.

a given decision problem. Note, however, that it may not be optimal in a strategic interaction between firm and customer. We thus abstract from incentives consumers may have to strategically adjust their attention allocation to the behavior of the firm.

⁷This might be the largest conceptual difference from other models of salience and limited attention. In our model, the reason why the weightings m_i vary across attributes is not that differences between alternatives are perceived as being larger or smaller than they actually are. In contrast, the weightings m_i express the extent to which differences between alternatives are appreciated in the decision process.

Assumption (b) captures the property that the salience of a dimension increases in the extent to which the alternatives differ in the particular attribute dimension. Hence, the consumer’s attention is responsive to the set of available alternatives. Therefore, attention is neither fixed nor random but depends on the context. In particular, we will assume that attention is a function of the total dispersion in an attribute. This is in line with models like Kőszegi and Szeidl (2011). The general intuition of a decision-maker neglecting small differences between options can already be found in Rubinstein (1988). Rubinstein argues that decisions as observed in the well-known Allais paradox can be explained by decision-makers treating similar probabilities or payoffs as equal.⁸

Assumption (c) describes the idea that the DM finds it increasingly difficult to consider an attribute when the complexity of the problem increases. This assumption implies that there is an attention hierarchy as it matters which attributes are considered “first”. Attributes that are considered earlier are more likely to be taken into account. In addition, we assume that this increased complexity cost is reflected in a lower attention weight given that the attribute is in fact considered. Thus, differences in attributes that are considered earlier are appreciated to a greater extent. Put shortly, each attribute has a rank in the attention hierarchy and a higher rank means that the attribute receives more attention.⁹

Note that this assumption ensures that all qualities that are considered by the DM are allocated a different attention parameter ($\forall m_i > 0, m_j > 0 : m_i \neq m_j$). This is important because it implies that behavior is indeed distorted by limited attention. To see this, recall from the functional form of the decision utility (1) that if attention would be uniformly dampened (e.g., with $m_i = 0.5 \forall i$), decision utility would just be a uniform transformation of experienced utility. In this case, a decision based on experienced utility is always the same as one based on decision utility. For limited attention to have a behavioral effect we need at least two attributes i and j which are allocated different levels of attention, i.e. $m_i \neq m_j$. Our assumptions make sure that any two attributes which are considered receive a different level of attention.

⁸Empirical results of Chetty, Looney, and Kroft (2009) indicate that consumers indeed often tend to pay higher attention to more dispersed attributes. They report that consumers are more sensitive to changes in the excise tax, which is included in the posted price, than to changes in sales taxes, which are added at the register. As the posted price is larger than the additional tax that is added at the register, the two alternatives “buying”/“not buying” differ more in the posted price dimension than they differ in the additional tax-dimension. Thus the more salient dimension seems to be the more dispersed one.

⁹See the Appendix for a further discussion of this assumption.

To construct an attention hierarchy which satisfies both (a) and (b) let μ_i measure the product of the valuation of attribute i and the maximal difference in attribute i between any two alternatives in the choice set:

$$\mu_i = v_i \left(\max_{a \in A} x_i^a - \min_{a \in A} x_i^a \right).$$

μ_i is the maximal difference in experienced utility a DM faces in dimension i in any binary comparison of two alternatives in the choice set. We will assume that an attribute's rank $r(i)$ in the attention hierarchy will be determined by the measure μ_i . Let $r : i \rightarrow \{1, \dots, n\}$ be the function that assigns an attention rank to each attribute such that

$$\begin{aligned} \mu_i > \mu_{i'} &\Rightarrow r(i) < r(i') \\ \mu_i = \mu_{i'} \text{ and } v_i > v_{i'} &\Rightarrow r(i) < r(i'), \end{aligned}$$

where $r(i)$ denotes the attention rank of the attribute i .¹⁰ Having determined the attention order we now turn to a cardinal measure of attention based on the attention hierarchy. Let m_i denote the attention paid to attribute i and define:¹¹

$$m_i = \max \left\{ 0, 1 - \kappa_{r(i)} / \mu_i \right\}.$$

The threshold $\kappa_{r(i)}$ is the minimum level of μ which an attribute with attention rank r needs to have in order to be taken into account. The threshold $\kappa_{r(i)}$ can be thought of as the cognitive cost of considering the r th dimension of a problem. This cognitive cost should reflect the fact that consumers have difficulties when having to choose between several product that are different in many attributes. We make the following assumptions with regard to the thresholds κ_r :

Assumption 1.

- (i) $\kappa_1 = 0$: *There is always one attribute that is fully considered.*
- (ii) $\kappa_r < \kappa_{r+1} \forall r \in \{1, \dots, n-1\}$: *The attention threshold is increasing with each additional attribute that is considered.*

¹⁰Note that $r(i)$ is well-defined since we assumed $v_i \neq v_j, \forall i \neq j$.

¹¹To let m_i always be well-defined let $m_i = 0$ if $\mu_i = 0$ and $\kappa_{r(i)} > 0$, and let $m_i = 1$ if $\mu_i = 0$ and $\kappa_{r(i)} = 0$.

Part (i) of the assumption ensures that there is always at least one attribute to which the DM pays full attention. The assumption captures that we model the DM to have difficulties with the complexity of comparing several attributes. If there is only a single dimension, the DM should face no problems. Part (ii) captures that the difficulty of considering an additional attribute increases with the number of attributes that are already considered.

Assumption 1 highlights the importance of the order of alternatives according to $r(i)$. First, the higher an attribute ranks in this order, the more likely it is attended to. If an attribute ranks low in the hierarchy, it is associated with a higher threshold κ , and hence it is more likely to be neglected. More specifically, if some attribute is taken into account, all attributes that rank higher must also be considered. Attention is a function of both *value* (v_i) - more valued attributes are more likely to be considered - and *contrast* - attributes in which alternatives differ greatly are more likely to be considered.

2.1 Comparison to other models of limited attention

The proposed model of attention allocation is conceptually very close to the one of Gabaix (2011). It takes over the idea of sparsity, meaning that some attributes are neglected by the DM if they are not “important enough”. A major difference to the attention process of Gabaix is that the threshold κ associated with an attribute i depends on an attribute’s *relative importance*, i.e. on how large its value and dispersion are compared to other attributes’ value and dispersion. While we retain the assumption that the vector of thresholds $(\kappa_1, \dots, \kappa_n)$ is exogenous, we impose a structure that we deem appropriate to reflect the notion of complexity costs. First, the decision-making process should not be distorted if the problem is not complex. We ensure this property by assuming $\kappa_1 = 0$. Second, we argue that an increase in complexity should be reflected by an increased difficulty to consider more and more dimensions of a problem. This motivates our assumption of increasing thresholds κ_r .

A second difference is the distinction between the salience, captured by m_i , and the value of an attribute, v_i . This distinction might seem superfluous at first glance as neither can be observed in isolation. Yet this simple distinction spares us the need to normalize our parameters to make salience independent of scaling. We find this feature desirable as we are concerned that some

of the behavioral implications which Gabaix derives are based heavily on this rescaling. In that sense certain behavior is predicted to occur not because people are inattentive but because this inattention is argued to be scale-independent.

The model we use shares a central property with the model of focus-dependent utility of Kőszegi and Szeidl (2011). They posit that the weighting of an attribute (m_i) is a strictly increasing function of the maximal utility difference associated with that attribute (μ_i). Our model shares this feature conditional on an attribute being considered, but not fully considered ($0 < m_i < 1$). Thus, a major difference is that our model features neglect. Kőszegi and Szeidl (2011) apply their model to a monopolist's problem to design a single product. They find an incentive to concentrate value of a product in one attribute and an incentive to split the price (and thus the disadvantage of a purchase) into multiple items. We are concerned that the first result is not an effect of the assumed focus-weighting alone but also depends strongly on the assumed cost function. In contrast, in our model there is an incentive to concentrate value that is entirely due to the weighting of attributes, i.e. in our context inattention. The reason is that if customers tend to neglect attributes of little importance, firms have an incentive to concentrate on few attributes in which they excel. The incentive to split the price into several items is also present in our model and is further strengthened by the consumers' tendency to neglect.

Bordalo, Gennaioli, and Shleifer (2010) present a model of salience-based choice. In contrast to our model, the modeling of salience is inspired by observations on perception (and not cognition) and assumes choice to be the result of multiple binary comparisons. More technically, the weighting does not only depend on the available choice set but also on the current binary comparison. Thus an attribute's salience in an alternative may be quite different depending on which two alternatives are currently examined.

3 Optimal Product Design of a Monopolist

In this section we investigate the problem of a monopolist who wants to design a product intended for sale to customers with the described attention process. What is central to our analysis is that attention is a function of the choices that are available. This implies that a firm is able to

manipulate attention as it can influence the choice set that a consumer faces. We will start by looking at the case in which the firm is restricted to offer a single product. In doing so, we begin by restricting the monopolist to a product with a single quality. We proceed by asking whether the monopolist would like to improve this product by introducing additional qualities. Thereby we derive the optimal design of the product. We then discuss under what conditions and in which way a firm may profit from introducing further products, and how the optimal product line is designed.

Suppose there are n qualities (save the price) that a product can have. Together with the price a product thus features $n + 1$ attributes. For now suppose that the level of each quality can take any non-negative real value: $x_i \in \mathbb{R}_+$, $\forall i = 1, \dots, n$. There is one attribute x_p which denotes the wealth of the decision maker. W.l.o.g. we normalize initial wealth to zero such that $x_p^a \leq 0$ means that alternative a is associated with a price of $P = -x_p^a$. Before we turn to the optimal design problem define a *null good* as an alternative with $x_i^0 = 0 \forall i = 1, \dots, n$.

A monopolist seeks to design a product that maximizes profits subject to the customer's willingness to buy it. To ensure that the customer is willing to purchase the good, the decision utility of the good (alternative a) must be weakly higher than the decision utility of abstaining from the purchase (alternative b). Note that alternative b is equivalent to a null good that is free of charge. Therefore, not buying is associated with a decision utility (and experienced utility) of zero. Part (i) of assumption 1 then implies that the monopolist cannot extract a positive profit by selling a null good at a positive price. Thus the product the monopolist designs must actually feature some positive qualities.

Let the costs of producing quality level x_i be $c(x_i) = \frac{1}{2}c_i x_i^2$. The monopolist then maximizes his profit subject to the decision utility of alternative a being nonnegative:

$$\begin{aligned} \max_{P, x_i} \quad & P - \frac{1}{2} \sum_{i=1}^n c_i x_i^2, \\ \text{subject to} \quad & \tilde{u}(a) \geq 0. \end{aligned}$$

First, let us look at the case in which a monopolist employs a single quality to equip his product with. In this case the set of feasible prices is constrained by the fact that the price must rank

second in the attention hierarchy: $v_i x_i \geq v_p P \Rightarrow r(i) = 1, r(p) = 2$. Otherwise the decision utility of buying the product would be strictly negative. This yields a maximum price $P^* = v_i x_i / v_p$. It is easy to calculate that the profit maximizing choice of the quality level x_i is

$$x_i^* = \frac{v_i}{v_p c_i}, \quad (2)$$

while the resulting profit is

$$\Pi^* = \frac{v_i^2}{2c_i(v_p)^2}. \quad (3)$$

Let us define the profitability measure π_i :

$$\pi_i \equiv \frac{v_i^2}{2c_i(v_p)^2}, \quad \forall i = 1, \dots, n.$$

The value π_i is the profit that a monopolist can extract from producing a product that features only quality i . We will refer to it as the *profitability of quality i* as it also denotes the maximum additional profit a monopolist could make by adding this quality to his product under full attention. It is straightforward to see that a monopolist who is confined to produce only a single-quality product will choose the quality with the largest π_i . It is interesting to compare this result to the one under full attention (full attention is equal to the case $\kappa_r = 0, \forall r$). First, note that if only one quality is produced, price, product design, and profit are equal to the solution under full attention. The sole difference is that under limited attention the decision utility of purchasing the product is strictly positive. While the price extracts all the value created by the quality, it is not perceived to do so (because it is not fully taken into account). Still it is not possible for the monopolist to exploit this and increase the price further. If the price was increased slightly above the monopoly price, the decision value created by the quality would drop due to a reallocation of attention from quality to price, thereby yielding negative decision utility and preventing the purchase altogether.

Now suppose the monopolist contemplates to introduce a second quality. This is only profitable if it allows to increase the price sufficiently to cover the cost of producing an additional quality. Until now the first quality was designed to rank higher in the attention hierarchy, i.e. $v_i x_i \geq v_p P$. For the optimal product this expression was fulfilled with equality. It follows that any price increase

above the level of the one-quality good moves the price to the top of the attention hierarchy. In other words, as soon as a product has at least two qualities, the price will rank first in the attention hierarchy. This shift in the attention rank lowers the decision utility of the first quality that the product already features and thus also lowers the profit that the monopolist can extract from the first quality. Therefore, the profit that can be extracted from the second quality must be sufficiently high to cover the decrease in profits that is caused by the downgrade in the attention rank of the first quality.

As before, the maximal price the monopolist can charge is given by the non-negativity constraint on the decision value of the product. Thus the monopolist will charge a price

$$P^* = \frac{1}{v_p} (m_i v_i x_i + m_j v_j x_j).$$

The optimal levels of the two qualities are:¹²

$$x_i = \begin{cases} \frac{v_i}{v_p c_i} & \text{if } m_i > 0, \\ 0 & \text{otherwise,} \end{cases}$$

$$x_j = \begin{cases} \frac{v_j}{v_p c_j} & \text{if } m_j > 0, \\ 0 & \text{otherwise.} \end{cases}$$

It is straightforward that the introduction of an additional quality is only profitable if both qualities are considered by the customer. But even if this is the case, the resulting profit may fall below the maximal profit of a product with a single quality. This is not because the new quality distracts attention from the quality the product was already equipped with. It may be the case that the new attribute ranks above the old attribute in the attention hierarchy. However, that effect does not have a negative impact on profits. Instead, the potential decrease in profits results from the fact that the price increase necessary to make the introduction of a new quality worthwhile distracts attention from the qualities and attracts attention to the price. Formally, the profit of the product

¹²The monopolist maximizes his profits by maximizing $\Pi = P^* - \sum_i c(x_i)$. To solve this recall the functional form of the attention parameter: $m_i = \max\{0, 1 - \kappa_{r(i)}/\mu_i\} = 1 - \kappa_{r(i)}/(x_i - 0)$.

with two qualities (if both of them are considered) is given by

$$\Pi^* = \pi_i + \pi_j - \kappa_2/v_p - \kappa_3/v_p. \quad (4)$$

Thus, the introduction of a second quality, say j , is profitable if and only if

$$\pi_j \geq \frac{1}{v_p}(\kappa_2 + \kappa_3). \quad (5)$$

From this profit expression it becomes clear that the firm is not benefiting from limited consumer attention. If consumers were fully rational, the firm could make a profit of $\Pi = \pi_i + \pi_j$. The fact that the firm makes lower profits if consumers have limited attention stems from consumers not fully appreciating the qualities of the product. Therefore, consumers with limited attention have a lower willingness to pay than rational consumers.

Now compare the profits made by producing a good with a single quality (3) to the profit made by producing a good with two qualities (5). The price increase that is made possible by the introduction of the second quality must be sufficient to cover both the physical cost of production and the cognitive cost of consideration of the new quality, and, in addition, the increase in cognitive cost associated with the consideration of the first attribute. It is again straightforward to see that, if confined to at most two qualities, the monopolist would choose the two qualities with the highest profitability π_i, π_j .

In the following we try to exemplify our results with small numerical examples. These examples seek to illustrate the mechanisms that drive our results. Let us first consider an example to illustrate the obstacles of introducing a second quality.

Example 1. *Suppose there are only two qualities with values $v_1 = 10$ and $v_2 = 9$. We normalize the value of money to $v_p = 1$. Initially, assume there are no cost differences between the two characteristics, $c_1 = c_2 = 3$. Finally, assume the cognition costs are $\kappa_r = 2^{r-1} - 1$, yielding $\kappa_1 = 0, \kappa_2 = 1$, and $\kappa_3 = 3$.*

In this setting it is optimal for the monopolist to equip his product with both qualities at the levels $x_1^ = 3.33$ and $x_2^* = 3$. This yields an optimal price of $p = 56.33$ and a profit of $\Pi = 26.17$. If the monopolist would offer a product with only one attribute, profits would be $\Pi = 16.67$ when*

using quality 1 and $\Pi = 13.5$ when using quality 2. Hence, the monopolist will offer a product with both qualities.

Now suppose that the second quality is more costly to produce with $c_2 = 15$. If the monopolist produces both qualities, the profit is $\Pi = 15.37$ which is less than the profit from equipping the product with only quality 1. Note that this is not the result of the customer not attending to quality 2. If a good with both attributes is produced, the optimal level of quality 2 would be $x_2^* = 0.6$ which yields positive attention factor $m_2 = 0.44 > 0$. However, the profitability of the second quality $\pi_2 = 2.7$ is not sufficient to cover the decrease in willingness to pay that is induced by the cognitive costs of attending the second quality $\kappa_3 = 3$ and the increase in cognitive costs associated with quality 1 $\kappa_2 - \kappa_1 = \kappa_2 = 1$. Thus for a low profitability of attribute 2 it is optimal to equip the product with only attribute 1.

Now assume that a product can be equipped with more than two qualities. Suppose we have derived the optimal design of a product that features two qualities. As we have argued before, this implies that the price is ranking first in the attention hierarchy. Hence, when considering adding a third attribute to the product, the necessary price increase will no longer be associated with a rise in attention allocated to the price since the price already ranks first in the attention hierarchy and thus already receives full attention. Thus, additional qualities will be introduced (naturally in the order of their profitability) if and only if their profitability is sufficient to cover the cognitive costs of considering an additional attribute. Hence, quality i is introduced if $\pi_i \geq \frac{1}{v_p} \kappa_{|\mathcal{I}|+1}$, where \mathcal{I} denotes the set of qualities that the product features (after the introduction of i).

Denote by $\pi^{(t)}$ the t -th most profitable attribute (ties may be broken according to v_i). Then we find the following result.

Proposition 1.

If a monopolist intends to supply a single good to the market with costs $c(x_i) = \frac{1}{2}c_i x_i^2$, the optimal design will feature:

- (i) a single quality i : $\pi_i = \pi^{(1)}$ if and only if $\pi^{(2)} < \frac{1}{v_p}(\kappa_2 + \kappa_3)$,*
- (ii) a second quality i : $\pi_i = \pi^{(2)}$ if and only if $\pi^{(2)} > \frac{1}{v_p}(\kappa_2 + \kappa_3)$,*
- (iii) and all further qualities i for which $\pi_i = \pi^{(t)} > \frac{1}{v_p} \kappa_{t+1}$.*

All qualities $i \in \mathcal{I}$ that the product features are produced at a level $x_i = v_i/(v_p c_i)$.

If the optimal good features only one characteristic, the price must rank second in salience, with $P = (v_i x_i)/v_p$, and is thus not getting full attention ($m_p < 1$). In this case, the decision utility of the product is strictly positive, i.e. $\tilde{u}(a) > 0$. If the optimal good features several characteristics, the price ranks first in salience ($m_p = 1$). The optimal price extracts the total decision value of the qualities, i.e. $\tilde{u}(a) = 0$. This yields a product price of:

$$P = \frac{1}{v_p} \sum_{i \in \mathcal{I}} m_i v_i x_i.$$

A couple of things are noteworthy. Given the product features a particular quality, its optimal level is the same as the one under full attention. This is due to two effects resulting from limited attention. First, limited attention reduces a firm's incentives to invest in quality as the created value is not fully taken into account by the customer. Second, the endogeneity of attention increases a firm's incentives to invest in quality as any increase in quality increases the attention paid to that attribute and thus the decision value of any unit of quality already invested.¹³ In our model, these two effects cancel out each other perfectly. The maximization problem of the monopolist then yields the same level of product quality as in the case with unlimited attention for those qualities which are considered. However, the product generally features less qualities than under full attention and the price is lower if the product features more than a single quality. Note here that the exact canceling of the above-mentioned effects hinges on our particular assumptions about the cost of production and on the specific modeling of limited attention, in particular our use of the range $(\max_{a \in A} x_i^a - \min_{a \in A} x_i^a)$ as a measure of attribute dispersion. Yet, we claim that the second effect prevails as long as one retains the assumption of a higher attribute dispersion attracting attention. As this second effect compensates the first, limited attention does not necessarily imply the production of lower quality levels.

To motivate our next section recall the profits (4) that the firm makes when it offers of product that features more than one quality. It holds that as soon as the product features more than one quality, the monopolist cannot extract the whole experienced utility that the product yields to the customer. This is because consumers decide on the basis of decision utility. In the case of multiple

¹³Note that the firm's incentive to invest in quality in order to increase attention is conditional on the attribute being considered. If the attribute is not considered, small changes in quality might be insufficient in order to lift attention to a positive level.

qualities the price receives full attention while all qualities are allocated less than full attention. Yet, the attention that is allocated to the qualities depends on the available choice set. That is what motivates the next section. We will analyze a way in which the monopolist can increase the customer's attention and thus the decision value of the product that is sold to consumers without changing the product itself.

4 Optimal Product Design with Introduction of a Bait Good

Recall that the consumer attention is a function of both the valuation v_i and the dispersion within the choice set ($\max x_i - \min x_i$). A firm might not be able to change the valuation v_i of an attribute, but it can manipulate the dispersion. If the firm can produce several goods, it might have an incentive to produce goods that have the sole purpose of increasing the dispersion of attributes, thereby making these attributes more salient. Note that the quality levels are bounded below by the option of not buying ($\min x_i = 0$). This implies that in order to increase dispersion, products that are designed to manipulate attention must feature high quality levels. This will increase dispersion by increasing $\max x_i$. Yet, if these products are not intended for sale, they must not be attractive to consumers. This could for example be achieved by a very high price. In the following, we will use the term *bait good* for those goods that have the sole purpose of manipulating consumer attention. Because these bait goods are designed to be unattractive, consumers still buy the main good which we will henceforth call *primary good*.¹⁴ Following this argument, we will now analyze how such a bait good is optimally designed.

In the analysis we will assume that the firm incurs no cost of designing and producing the bait good. This is obviously not realistic in most circumstances. We maintain this assumption to concentrate on two questions. First, is it possible to increase the willingness-to-pay of consumers for the primary good, and thus the profit made from its sale, by simultaneously offering a bait good without distracting demand from the primary good? Second, if such an attention manipulation is possible, how is it done optimally? In other words, how is a product solely intended for attention manipulation optimally designed?

¹⁴Obviously, the primary good still has to yield positive decision utility to the consumer in order to be bought.

There are two further arguments for neglecting the costs of the bait good. First, if the bait good is never actually sold to the customer, it only needs a single item of the bait good that can be (unsuccessfully) offered to each of a large number of customers. If, as we will show, the bait good increases profits, the additional profits reaped from each customer may in sum be sufficient to cover the cost of bait good production. A second avenue to accommodate the cost of producing the bait good is to allow for customer heterogeneity in the value of money. As we will argue in the discussion section, a product designed for a richer customer segment is a prime candidate to serve as a bait good for a poorer customer segment. In this way, the bait good is produced both for sale (to a rich customer segment) as well as for the manipulation of attention (of a poorer customer segment).

There are two ways how a manipulation of attention could work. First, the firm could try to draw more attention to the qualities of the primary good without changing the attention hierarchy. Second, the firm could try to manipulate the attention hierarchy to redistribute attention between qualities. As mentioned before, we will show that the firm has an incentive to produce premium quality goods. In reality, maximum quality levels are of course bounded at some level. For example, there are technological limitations on the horsepower that a car can have. Therefore, we will now assume technical frontiers for each quality. Formally, all $x_i \in [0, \bar{x}_i]$ with $\bar{x}_i \in \mathbb{R}^+$.¹⁵

Now let us look at the introduction of a bait good that is designed to attract attention to the qualities of the primary good, but without changing the attention hierarchy of the qualities embodied in the primary good. Note first that if the monopolist is confined to employing only a single quality, there is nothing to be gained from manipulating attention. This is because the primary good is already designed such that the quality employed ranks first in the attention hierarchy thus receiving full attention. A bait good will thus derive its value through increasing the attention paid to qualities which are not fully considered (or neglected altogether). We will first show that whenever the primary good features more than one quality at levels below their respective technological frontiers, it is profitable to employ a bait good in order to increase attention and thereby increase consumers' willingness to pay.

¹⁵Note that without such technological frontiers the monopolist would be able to induce full attention in the limit. But even without this technical issue, it seems realistic to assume a cap on the technological possibilities that the monopolist has at his disposal. In addition, it allows us to make some predictions concerning the impact of these technological barriers.

Proposition 2. *Suppose a monopolist's profit-maximizing product design features at least two qualities: $|\mathcal{I}| \geq 2$. If at least one of these qualities is not produced at the highest technologically feasible level, the monopolist can strictly increase profits by using a bait good.*

Proof. Consider a monopolist whose product features $|\mathcal{I}| \geq 2$ qualities. Further suppose that the optimal level of one of these qualities is below its technological frontier, i.e. $\exists i \in \mathcal{I} : x_i = \frac{v_i}{v_p c_i} < \bar{x}_i$. Let the monopolist offer a second product, called bait good, which is equipped with all the qualities that the primary good features at their respective level. The qualities of the bait good are denoted as x_i^b . The bait good's quality i is increased slightly to $x_i^b = \frac{v_i}{v_p c_i} + \epsilon$ for some small $\epsilon > 0$. Such a small increase will not affect the attention order except for cases in which attribute i is already at the threshold for reaching a higher rank.¹⁶ Still, the attention paid to attribute i increases to $m_i = 1 - \kappa_{r(i)}/(v_i x_i^b)$. The bait good's price is set appropriately high to make its purchase unattractive. This high bait good prices has no impact on consumer attention since, due to Proposition 1, the price already ranks first in the attention hierarchy with $m_p = 1$. The maximal price the monopolist can charge for the primary good increases to

$$P^* = \frac{1}{v_p} \left[\sum_{j \in \mathcal{I}} m_j v_j x_j \right] = \frac{1}{v_p} \left[\sum_{j \in \mathcal{I} \setminus i} (v_j x_j - \kappa_{r(j)}) + \left(v_i - \frac{\kappa_{r(i)}}{x_i^b} \right) x_i \right].$$

Given this optimal price, the optimal levels of the qualities are

$$\begin{aligned} x_j^* &= \frac{v_j}{v_p c_j} \quad \forall j \in \mathcal{I} \setminus i \\ x_i^* &= \frac{1}{v_p c_i} \left(v_i - \kappa_{r(i)} \frac{1}{x_i^b} \right). \end{aligned}$$

The maximal profit a monopolist can make with the help of the bait good then is

$$\Pi^* = \sum_{j \in \mathcal{I} \setminus i} \left(\pi_j - \frac{1}{v_p} \kappa_{r(i)} \right) + \pi_i m_i^2.$$

Note how the use of a bait good changes the profit that a monopolist can make out of feature i . In general, if the monopolist uses a bait good that features all the qualities of the primary good

¹⁶Even in this case, it is easy to verify that employing the bait good is profitable.

(but with higher levels), the maximal profit is given by

$$\Pi^* = \sum_{i \in \mathcal{I}} \pi_i m_i^2. \quad (6)$$

To show that the monopolist increases his profit with the use of a bait good we need to show that $\pi_i m_i^2 > \pi_i - \kappa_{r(i)}/v_p$. Note that

$$\pi_i m_i^2 = \pi_i - 2\pi_i \frac{\kappa_{r(i)}}{v_i x_i^b} + \pi_i \left(\frac{\kappa_{r(i)}}{v_i x_i^b} \right)^2.$$

Then, it suffices to show that

$$\begin{aligned} \frac{\kappa_{r(i)}}{v_p} &> 2\pi_i \frac{\kappa_{r(i)}}{v_i x_i^b} \\ \Leftrightarrow x_i^b &> \frac{v_i}{v_p c_i}, \end{aligned}$$

which is true by construction of x_i^b . □

The bait good thus increases the profit that can be made from the sale of the primary good while keeping the attention hierarchy among attributes intact. Note here that we have just shown that a bait good can increase profits under fairly simple conditions. We have not yet looked at the profit maximizing design of the bait good. In fact, it might be the case that the firm can make even higher profits if the bait good induces a shift in the attention hierarchy. We will investigate this case later in this section. Now let us revisit our initial example.

Example 2. *First, let us consider the case where $v_1 = 10$, $v_2 = 9$, $v_p = 1$, $\kappa_1 = 0$, $\kappa_2 = 1$, $\kappa_3 = 3$, and $c_1 = c_2 = 3$ as in example 1. In addition, let us assume technological frontiers of $\bar{x}_1 = \bar{x}_2 = 6$. If the monopolist designs a bait good with qualities $x_1^b = \bar{x}_1$, $x_2^b = \bar{x}_2$ and a prohibitively high price, the customers attention for both qualities of the primary good will increase. The high price ensures that the consumer is not being tempted to buy the bait good itself. Given this bait good, the monopolist's optimal design of the primary good features both qualities at levels $x_1^* = 3.28$ and $x_2^* = 2.83$. This yields a profit of $\Pi = 28.16$ that exceeds the maximal profit of $\Pi = 26.17$ that the firm made without the bait good.*

Also note that Proposition 2 is only a sufficient condition for the profitability of the bait good. There are two other cases in which the utilization of a bait good increases firm profits. The first case concerns circumstances in which, without usage of a bait good, it was optimal to produce a primary product with only one quality. Then the introduction of the bait good may allow the firm to profitably introduce further qualities.

Example 3. *Let us reconsider the second case in example 1 in which the higher costs of the second quality $c_2 = 15$ made its introduction unprofitable. Using the same bait good as above in example 2, the monopolist can make a profit of $\Pi = 18.52$ by equipping the primary good with both qualities. This exceeds the profit that the firm can make by offering a primary good with only a single quality. The use of a bait good can thus make the introduction of qualities profitable that are unprofitable without the bait good.*

The second case in which the bait good can increase profits besides the cases already mentioned in Proposition 2 concerns circumstances in which the primary product is bounded by the technological constraints. In that case consumer attention cannot be increased by the usage of a bait good. Recall at this point that without a bait good, there is an incentive to invest in costly quality because quality has an impact on consumer attention. Now this manipulation of consumer attention can be costlessly attained by the bait good. Hence, if initially the primary product was produced at the technological constraint, it might be profitable to introduce a bait good that is also at the technological constraints and lower the qualities of the primary good.

In addition to merely increasing the attention paid to each attribute of the primary good, a bait good can be employed to shift the attention order, e.g. to push a more profitable characteristic to a more salient position. We will now investigate under what conditions this is profitable. Suppose for simplicity that the primary good features two qualities $i = 1, 2$ and assume that $m_1 > m_2$. Further assume that the bait good in this setup features qualities $x_1^b = \bar{x}_1$ and $x_2^b = v_1\bar{x}_1/v_2 < \bar{x}_2$. Thus the bait good is designed in such a way that the primary good yields maximal profit given that the first quality ranks higher in the attention order than the second quality. Increasing x_2^b above $v_1\bar{x}_1/v_2$ would change the attention order: the first quality would become less, the second quality more salient. With the help of equation (6) it is easy to see that this change is profitable

if the additional profit made from quality 2 exceeds the reduction in profits from quality 1:

$$\pi_2 (m_{2,new}^2 - m_{2,old}^2) > \pi_1 (m_{1,old}^2 - m_{1,new}^2).$$

This equation is likely to hold if (a) quality 2 is more profitable than quality 1, and/or (b) the technology frontier \bar{x}_2 of quality 2 greatly exceeds the restriction $v_1\bar{x}_1/v_2$ that the bait good would have to satisfy in order to retain the old attention order.

Let us consider an example that highlights that it is not always optimal to employ a bait good that features all qualities at their technological frontier, i.e. it is not always optimal to offer the best product as bait good that is technologically feasible.

Example 4 (Profitability). *This part seeks to illustrate how the bait good can be employed to draw attention to more profitable qualities. Suppose customer valuations are $v_1 = 10$, $v_2 = 9$, and $v_p = 1$ as before, but costs are $c_1 = 6$ and $c_2 = 3$. The profitability values $\pi_1 = 8.33$ and $\pi_2 = 13.5$ show that the higher-valued quality is less profitable. Suppose the bait good was constructed as in examples 2 and 3 by setting $x_1^b = \bar{x}_1 = 6$ and $x_2^b = \bar{x}_2 = 6$. Then the less profitable quality would rank higher (at rank 2, behind the price) than the more profitable quality (rank 3). This yields a profit of $\Pi = 20.10$. From the profit formula for the case with bait goods $\Pi = \sum_{i \in \mathcal{I}} \pi_i m_i^2$ it follows that a monopolist may profit from manipulating the attention hierarchy such that the most profitable quality gets the most attention. Suppose x_1^b was lowered to 5. This would make quality 2 rank second in the attention order, and quality 1 rank third, yielding a profit of $\pi = 20.36$.¹⁷ This example illustrates that the bait good cannot only be used to attract attention, but also to distract attention. This is attractive if a firm wants to distract customers' attention from a more valuable, but less profitable quality to a less valuable, but more profitable quality.*

Example 4 highlights that firms usually have an incentive to make the most profitable qualities most salient. However, this does not imply that the optimal design will always make the most profitable qualities the most salient. We will now give an example in which the firm prefers to rank a high-technology attribute higher than a high-profitability attribute. Consider the following example.

¹⁷Actually, it would be sufficient (and thus optimal) to lower x_1^b slightly below $(v_2\bar{x}_2)/v_1 = 5.4$. If, e.g., $x_1^b = 5.3$ was chosen, the profit would be $\Pi = 20.42$.

Example 5 (Technology differences). Assume $v_1 = 10$, $v_2 = 9$, $v_p = 1$, and $c_1 = c_2 = 3$ as in example 1. However, assume that $\bar{x}_1 = 6$ and $\bar{x}_2 = 60$. As in example 1, quality 1 is more valuable to the customer and more profitable to the firm than quality 2. It is optimal to set $x_1^b = \bar{x}_1 = 6$ irrespective of the attention rank that quality 1 is assigned in the end. Suppose it is assigned rank 2 (the price has rank 1) while quality 2 is assigned rank 3. Then x_2^b is restricted to the value $v_1\bar{x}_1/v_2 = 10 \cdot 6/9 = 6.67$, because otherwise quality 2 ranks higher than quality 1. The resulting maximal profit for the firm is $\Pi = 28.30$. If, however, the firm sets $x_1^b = \bar{x}_1 = 6$ and $x_2^b = \bar{x}_2 = 60$, the less profitable quality 2 will rank second in the attention hierarchy, while the more profitable quality 1 ranks third. This gives the firm a profit of $\Pi = 28.49$ which is an improvement. In this example, for all values of \bar{x}_2 above ≈ 12.5 it is optimal for the firm to make the higher-valued and more profitable quality less salient. It therefore illustrates that a firm may sometimes have an incentive to use a bait good to distract from a more profitable quality. This should, however, only occur in situations in which there are small differences in profitability but huge differences in the technological limitations associated with the two qualities.

The bait good is effective in attracting attention when it features qualities at a high level. When contemplating the optimal attention manipulation the firm faces the following problem. On the one hand, a firm has an incentive to let profitable qualities rank high in the attention order (see example 4). This allows to attract a lot of attention to these characteristics, which enables the firm to extract a larger profit. On the other hand, given some assigned ranking, a firm wants to maximize attention paid to each quality. This second incentive may conflict with the first. Assigning a particular rank to a quality imposes a restriction. In order for this quality to keep the assigned rank, the extent of attention that can be attracted to all qualities at lower ranks is limited. The firm may thus not be able to exploit the technological boundaries of lower-ranking attributes, which lowers the profit that can be extracted from these lower ranking attributes.

We can conclude that the optimal design of the bait good will take into account both the profitability of the qualities and the technological possibilities. Until now we have assumed that the firm can produce only one bait good. In the next section we will investigate under which circumstances it might be optimal to employ several bait goods.

5 Usage of several Bait Goods

Until now we have assumed that the monopolist can employ exactly one bait good. The purpose of the bait good is to increase attention for (some) qualities that are featured by the primary good. This increase in attention allows the firm to charge higher prices since it leads to an increase in the consumers' willingness to pay. In particular, the bait good increases attention of some attributes if it is equipped with extremely high levels of these quality. However, since the firm does not actually want to sell the bait good, it has to charge a very high price for it to make it less attractive to consumers than the primary good. Recall that this implied that the price kept the highest rank in the attention hierarchy. Since the price already ranked first under the optimal design of a single product (if $|\mathcal{I}| \geq 2$), thus being fully considered with $m_p = 1$, the high bait good price had no influence on the salience of the price dimension.

When the firm is allowed to produce more than one bait good it can use a different approach towards attention manipulation. Suppose the firm initially produces a primary good that features several qualities. One potential bait good that the firm could introduce is one that features a high level of the first quality and no additional qualities. This would increase the attention that consumers pay to the first quality. Note that increasing the salience of qualities could already be achieved with a single bait good discussed in the prior section. However, since the new bait good only features one quality (and the primary good features many), the bait good's price does not need to be very high in order to make its purchase unattractive. As we will show, the usage of several specialized bait goods can make it feasible to make some qualities even more salient than the price. This reduces the rank and thus the salience of the price. This might be profitable because if the price is less salient, the willingness to pay of consumers increases which allows the firm to raise the price of the primary good.

In the following, we assume w.l.o.g. $v_i > v_{i+1} \forall i = 1 \dots n - 1$. Consider the case in which the firm wants to make the most valuable quality 1 more salient than the price. When employing a specialized bait good, the firm has to meet several restrictions for the design of the bait good. First it must still hold that the bait good is less attractive than the primary good:

$$\tilde{u}^{primary} \geq \tilde{u}_1^{bait}. \quad (7)$$

In addition, the bait good should not increase the salience of the price, i.e. $p_1^{bait} \leq p^{primary}$. Since the bait good should be unattractive, we can focus on the case $p \equiv p_1^{bait} = p^{primary}$. For a bait good with quality 1 we can then write (7) as:

$$\begin{aligned} & \sum_{i=1}^n m_i v_i x_i^{primary} - m_p v_p p \geq m_1 v_1 x_1^{bait} - m_p v_p p \\ \Leftrightarrow & \sum_{i=1}^n m_i v_i x_i^{primary} \geq m_1 v_1 x_1^{bait}. \end{aligned} \quad (8)$$

As the objective is to push the price to a lower attention rank, the bait good must be designed such that quality 1 ranks higher than the price. So if a bait good features only quality 1, the bait good must suffice

$$v_1 x_1^{bait} \geq v_p p. \quad (9)$$

Combining (8) and (9), it must hold that:

$$\sum_{i=1}^n m_i v_i x_i^{primary} \geq m_1 v_1 x_1^{bait} \geq m_1 v_p p. \quad (10)$$

Note here that the price was initially at the highest attention rank. Hence, if quality 1 gets pushed in front of the price, it is now ranking first with $m_1 = 1$. Then (10) can be written as:

$$\sum_{i=1}^n m_i v_i x_i^{primary} \geq v_1 x_1^{bait} \geq v_p p. \quad (11)$$

By setting x_1^{bait} and p such that all three terms in (11) have equal value, the firm can charge a maximum price of:

$$p = \frac{1}{v_p} \sum_{i=1}^n m_i v_i x_i^{primary}.$$

This looks like the pricing formula in the case with only a single bait good. Note, however, that m_1 has increased relative to the case with only one bait good (because quality 1 now ranks one rank higher). Hence, by employing a specialized bait good the firm raised the price that it can charge for the primary good.

Until now we have described how attribute 1 can be pushed in front of the price in the ranking of salience. We can proceed by doing the same with qualities 2, 3, and so on. When quality 2 shall gain attention rank 2, the bait good for quality 2 must be designed such that it satisfies:

$$v_1 x_1^{bait} \geq v_2 x_2^{bait} \geq v_p p.$$

If this condition can be fulfilled, the usage of the second bait good increases m_2 because quality 2 would gain one rank in the attention hierarchy.

In principle this process can be continued for all other qualities. For all qualities that are more salient than the price it must hold that $v_i x_i^{bait} = v_{i+1} x_{i+1}^{bait}$. Note here that since $v_i > v_{i+1}$, each additional bait good must be produced to have a higher level of the quality on which it is specialized, i.e. $x_i^{bait} < x_{i+1}^{bait}$. This may not be feasible for all qualities due to the technological bounds \bar{x}_i . Hence, it may not be possible to make all qualities more salient than the price. However, even for qualities that rank below the price attention can be increased by using bait goods. Just as before, qualities that rank below the price can still be made more salient relative to the no-bait good case by producing a bait good that has a quality level for the given attributes that is above the quality level of the primary good. This procedure still increases attention for these qualities. This effect is basically the same as in the case with only one bait good.

Example 6 (Several bait goods). *Suppose that we have $v_1 = 10$, $v_2 = 9$, $v_p = 1$, $c_1 = 3$, and $c_2 = 3$. In addition, we assume that $\bar{x}_1 = \bar{x}_2 = 6$. We have seen before that without the usage of a bait good, the firm would make a profit of $\Pi = 26.17$. By using one single bait good that features both qualities the firm is able to increase its profit to $\Pi = 28.16$. The firm achieved this increase in profits by producing a bait good with $x_1^{bait} = \bar{x}_1 = 6$ and $x_2^{bait} = \bar{x}_2 = 6$. This leads to an increase in attention factors m_1 and m_2 , relative to the case without a bait good. This increase in the attention factors then increased the consumers' willingness to pay. However, since the bait good was not intended for sale (because it would be too costly to produce), the firm had to set a high bait good price which implied that the price remained in attention rank 1.*

If the firm can produce two bait goods, it turns out that it is easier to make them unattractive. This is basically because the primary good features two qualities and each bait good features only one quality. Let us consider the optimal bait good design. The firm wants to achieve that (some)

qualities are more salient than the price dimension. This can be achieved by producing a bait good that features a high level of its particular quality. However, the quality level of the bait good should not be too high because the bait good should still be less attractive than the primary good. Since the price is dependent on the resulting attention structure, we have to solve this problem recursively. This is necessary because the bait good design is dependent on the price, but the price is also dependent on the bait good design.

Solving this problem, we find that the optimal qualities of the primary good are $x_1^{\text{primary}} = 3.33$ and $x_2^{\text{primary}} = 2.83$ while the price is $p = 57.42$. The bait good that features quality 1 is produced with $x_1^{\text{bait}} = 5.74$ and $p^{\text{bait}} = p = 57.42$. This ensures that $v_1 x_1^{\text{bait}} \geq v_p p$ and hence quality 1 is more salient than the price. The firm would also like to make quality 2 more salient than the price. This would require $v_2 x_2^{\text{bait}} \geq v_p p$, yielding $x_2^{\text{bait}} \geq 6.38$. This is not feasible due to $\bar{x}_2 = 6$. Thus the firm is confined to offer a second specialized bait good at the highest technological level with $x_2^{\text{bait}} = 6$. Hence, quality 1 will rank first in the attention hierarchy, followed by the price and quality 2. This yields attention parameters of $m_1 = 1$, $m_p = 0.98$, and $m_2 = 0.94$. The resulting profit of the firm is $\Pi = 28.71$, which is higher than in the case with only a single bait good.

This example illustrates several effects that occur when using several bait goods. Note that the quality of the first bait good is restricted to $x_1^{\text{bait}} = 5.47$ in order to not be more attractive than the primary good. A similar restriction would be in place for all qualities that are intended to be made more salient than the price. For example, quality 2 would have to be no higher than $x_2^{\text{bait}} = 6.66$.¹⁸ Note that for all qualities that do not rank first in the attention hierarchy these restrictions limit the attention that can be drawn to these qualities. Suppose that the technological limits are very high (say $\bar{x}_2 = 50$). Then it might be the case that it is more profitable to employ a single bait good and exploit the technological boundaries. Although this implies that quality 2 receives a lower rank (3 instead of 2), it will still receive more attention at this lower rank as the bait good may exploit the technological boundary of 50 instead of being confined to 6.66. Hence, it will turn out that even if it is feasible to make a quality more salient than the price, it might not always be optimal to do so.

¹⁸A value $x_2^{\text{bait}} = 6.38$ is sufficient to increase the attention rank of quality 2. However, due to the increased attention, it is then optimal to increase the quality 2 of the primary good a bit. This increases the attractiveness of the primary good such that x_2^{bait} can be further increased. Solving this recursively, the maximum value of the second bait good in order to still be less attractive than the primary good is $x_2^{\text{bait}} = 6.66$.

To see this consider the outlined example with technological bounds $\bar{x}_1 = \bar{x}_2 = 50$. In this case the bait good levels that would be chosen in order to make attributes 1 and 2 more salient than the price would be $x_1^{bait} = 6$ and $x_2^{bait} = 6.66$. This would imply attention factors of $m_1 = 1$, $m_2 = 0.98$, $m_p = 0.95$, and a profit of $\Pi = 29.72$. If, however, the firm would use only one bait good, it would produce its bait goods to have maximum quality levels, i.e. $x_1^{bait} = x_2^{bait} = 50$.¹⁹ This would yield attention factors of $m_p = 1$, $m_1 = 0.998$, and $m_2 = 0.99$. In this case the profit of the firm would be $\Pi = 29.92$, which is higher than in the case with two specialized bait goods.

We conclude that the usage of several specialized bait goods is only profitable if the technological limits on quality levels are neither too low (making qualities more salient than the price allows only very small prices) nor too high (usage of a single bait good that exploits the technological limits is better than trying to distract attention from the price through specialized bait goods).

6 Discussion

This paper seeks to highlight how a firm may employ certain products to manipulate the attention process of its customers. There are a couple of issues and potential extensions of the model that are worth further discussion.

6.1 Heterogeneous Consumers

Until now we have assumed that consumers are homogeneous in both their preferences and their cognitive abilities. This assumption could be weakened in several ways. One interesting possibility is to allow customers to have different levels of wealth. For simplicity assume there are two groups of customers, very loosely labeled some are *poor* and some are *rich*. Assuming a diminishing marginal utility of money, let the first group assign higher importance to the money dimension (and thus the price) of a product: $v_p^{poor} > v_p^{rich}$. This will result in a firm designing different products to cater to both groups of customers. The rich group's product may feature more qualities than the poor group's product. It certainly features all the qualities the poor group's

¹⁹In this case it is irrelevant whether the firm produces one or two bait goods.

product has, but on a higher level (recall the solutions for the optimal levels of qualities). This makes the rich group’s product an ideal candidate to be a bait good for the poor group. It increases the attention paid to all the qualities that the product for the poor features and thereby increases the willingness to pay of the poor. In this way a firm may employ products designed for richer customer segments as a bait good for poorer customer segments. This might explain one of the examples we discussed in the introduction: the advertisement of expensive cars to an audience of which a majority is not able to afford it.

If customers differ in the cognitive constraints they face, e.g. because one group has to act under stronger time pressure, a firm may cater to these different groups with products differing in their degree of complexity, i.e. differing in the number of qualities they feature. The hurried customer segment is then offered a product with only a few essential features, while those customers with more time prefer more elaborate products.

Finally, customers may differ in their valuation for different qualities. As these valuations influence both an attribute’s profitability and the relative ease of attention attraction, a firm may have strong incentives to segment the customer population and design an appropriate product line for each segment separately. Still, analogous to the discussion of differing wealth levels, a firm may have an incentive to offer a product designed for one customer segment to a different segment despite that segment’s unwillingness to purchase the product. Offering a sports car to a family father may increase his willingness to pay for horse power despite his general focus on car safety.

6.2 Negative “Qualities”

So far we have assumed that all attributes (save the price) are valued by the customer ($v_i \geq 0 \forall i$). There are certainly some characteristics that a product may feature which customers dislike, e.g. the level of exhaust fumes of cars, or the level of sugar and trans fats in food. While the customer dislikes these characteristics and their presence reduces consumers’ willingness-to-pay, a reduction and/or replacement of these attributes may be costly to the firm. Employing bait goods can thus have further beneficial effects. In addition to attracting attention to positive characteristics, the firm could distract from negative characteristics. In addition, a firm might also have incentives to focus on reducing few negative characteristics a lot instead of reducing all

negative characteristics a little bit. A label “sugar free” or “low carb” may effectively distract from other negative characteristics, for example high levels of trans fats.

6.3 Attribute Dependence

We assume attributes to be independent. One might consider relaxing this condition to allow for complementarity and/or substitutability between attributes in production and/or consumption. In the case of dependencies in consumption, the importance of an attribute v_i then is a function of the level of complementary/substitute attributes. An attribute’s level may thus decrease the attention paid to a substitute attribute even if both attributes are several ranks apart in the attention hierarchy.

There is a complicating issue connected with allowing for dependencies in consumption between attributes. The question is whether the DM does take into account dependencies between all attributes or only between those that he takes into account? What if there is a dependency between an attribute that is considered and an attribute that is neglected? In addition, one might wonder why a DM seeks to decrease the complexity of a decision problem by neglecting some attributes but has no problems taking into account complex interactions between attributes.

7 Conclusion

We have introduced a new approach to model limited attention and applied it to the problem of optimal product design. The proposed attention heuristic fulfills several desirable properties that we think are realistic in real world markets.

Using this framework, we have shown that limited attention has far-reaching implications for product design and in general also for product lines. We have started with the case in which the monopolist can offer only a single product. In most cases firms produce products that have fewer attributes than they would have if consumers were fully rational. However, attributes that have positive quality may actually be produced at the same level as in a fully rational model. We find that a monopolist would actually prefer consumers to be fully rational. This is because the

consumers' willingness-to-pay is lower under limited attention since they do not fully appreciate all the qualities inherent in a product.

Since the monopolist profits from an increase in attention, he has an incentive to introduce goods of premium quality that are not intended for sale, but which increase the attention of the consumers. These bait goods can be used to increase attention, but they can also distract consumer attention from attributes that are less profitable to the monopolist. Furthermore, they can be used in such a way that the price appears less salient to consumers, which also increases the consumers' willingness to pay and thereby increases monopoly profits.

References

- Benartzi, S. and R. Thaler (2002). How Much is Investor Autonomy Worth? *The Journal of Finance* 57(4), 1593–1616.
- Bordalo, P. (2011). Choice-Set Effects and Salience Theory. Working Paper.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2010). Salience Theory of Choice Under Risk. Working Paper.
- Chetty, R., A. Looney, and K. Kroft (2009). Salience and Taxation: Theory and Evidence. *The American Economic Review* 99(4), 1145–1177.
- Eliasz, K. and R. Spiegel (2011). On the strategic use of attention grabbers. *Theoretical Economics* 6(1), 127–155.
- Gabaix, X. (2011). A Sparsity-Based Model of Bounded Rationality. Working Paper.
- Herne, K. (1997). Decoy alternatives in policy choices: Asymmetric domination and compromise effects. *European Journal of Political Economy* 13(3), 575–589.
- Huber, J., J. Payne, and C. Puto (1982). Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis. *The Journal of Consumer Research* 9(1), 90–98.
- Kahneman, D. (2000). Experienced Utility and Objective Happiness: A Moment-Based Approach. In D. Kahneman and A. Tversky (Eds.), *Choices, Values, and Frames*, pp. 673–692.

Cambridge University Press.

Kahneman, D. and A. Tversky (1984). Choices, Values, and Frames. *American Psychologist* 39(4), 341–350.

Kamenica, E. (2008). Contextual Inference in Markets: On the Informational Content of Product Lines. *The American Economic Review* 98(5), 2127–2149.

Kőszegi, B. and A. Szeidl (2011). A Model of Focusing in Economic Choice. Working Paper.

McFadden, D. (1999). Rationality for Economists? *Journal of Risk and Uncertainty* 19(1), 73–105.

Osborne, M. and A. Rubinstein (1998). Games with Procedurally Rational Players. *The American Economic Review* 88(4), 834–847.

Rubinstein, A. (1988). Similarity and Decision-making under Risk (Is There a Utility Theory Resolution to the Allais Paradox?). *Journal of Economic Theory* 46(1), 145–153.

Rubinstein, A. (1993). On Price Recognition and Computational Complexity in a Monopolistic Model. *Journal of Political Economy* 101(3), 473–484.

Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *The Journal of Consumer Research* 16(2), 158–174.

Simonson, I. and A. Tversky (1992). Choice in Context: Tradeoff Contrast and Extremeness Aversion. *Journal of Marketing Research* 29(3), 281–295.

Spiegler, R. (2006a). Competition over agents with boundedly rational expectations. *Theoretical Economics* 1(2), 207–231.

Spiegler, R. (2006b). The Market for Quacks. *The Review of Economic Studies* 73(4), 1113–1131.

Vikander, N. (2010). Targeted Advertising and Social Status. Working Paper.

Wernerfelt, B. (1995). A Rational Reconstruction of the Compromise Effect: Using Market Data to Infer Utilities. *Journal of Consumer Research* 21(4), 627–633.

8 Appendix: Derivation of an Optimal Attention Allocation under Cognitive Constraints

Suppose a decision maker faces the problem of choosing between a finite number of alternatives from the set A . Each alternative is described by a finite vector of attributes $i \in I$. Let x_i^a denote the extent to which alternative a features attribute i . The hedonic utility of each alternative $a \in A$ is expressed by

$$u(a) = \sum_i v_i x_i^a, \quad (12)$$

where v_i denotes the value the DM ascribes to an additional unit x_i of attribute $i \in I$. Thus the choice problem can be expressed by two vectors (A, v) . Suppose a decision-maker (DM) faces cognitive constraints such that he incurs cognitive costs whenever he faces a choice between multi-attribute alternatives. i.e. $|I| \geq 2$. The DM faces no information problem, he perfectly knows the values $(x_i^a)_{a \in A}$. He, however, faces problems whenever he needs to integrate this information in order to make a choice. He thus imperfectly considers or takes into account the information, and thus evaluates each alternative by his decision utility given by

$$\tilde{u}(a) = \sum_i [m_i v_i x_i^a + (1 - m_i) v_i x_i^d], \quad (13)$$

where $m_i \in [0, 1]$ can be thought of as an attention parameter.²⁰ $m_i = 1$ denotes full attention, while $m_i = 0$ denotes complete neglect. When neglecting the information in one dimension/attribute, the DM ascribes some value x_i^d to each alternative. Depending on the assumptions one wants to make, this default value may differ. If the DM is “on average right”, one may choose $x_i^d = \bar{x}_i$ where \bar{x}_i is the average value of x_i across the available alternatives. If the DM is pessimistic, $x_i^d = \min_{a \in A} x_i^a$ might be a good assumption. If the DM has some default alternative, $x_i^d = x_i^{default}$ could be reasonable. Regardless of these assumptions, if the DM neglects a dimension, i.e. if $m_i = 0 \forall i \in I$, he is not able to discriminate between alternatives along this dimension.²¹

²⁰Note here that the second term $(1 - m_i)v_i x_i^d$ only enters as a constant in the utility function. Hence, this utility function (13) yields the same choice behavior as the utility function (1) that was used in the main text.

²¹The exact assumptions on x_i^d are irrelevant for the behavior of the DM. If a dimension is neglected, a constant $v_i x_i^d$ is added to the decision utility of all alternatives. Of course, any constant added to a utility function has no

Now, let's look at the error from imperfectly considering the dimensions:

$$\sum_{a \in A} \left[\sum_i (v_i x_i^a - m_i v_i x_i^a - (1 - m_i) v_i x_i^d) \right] = \sum_i \left[(1 - m_i) v_i \sum_{a \in A} (x_i^a - x_i^d) \right].$$

An optimal attention allocation will weigh losses from an erroneous representation against losses from incurred cognition costs.²² Remember that we seek to model cognitive costs associated with complexity. It should thus be straightforward to assume that considering a single dimension is costless, as there is no complexity involved. Considering a second dimension involves the need to make a first trade-off consideration and should thus be associated with some positive cognition costs. Taking into account additional dimensions should become increasingly costly as the number of trade-off considerations that need to be made rises exponentially. It thus matters which dimensions are considered “first”. Denote by $r : I \rightarrow \{1, \dots, n\}$ the order in which the attributes are considered. We will henceforth refer to it as the attention hierarchy. Given some place in the hierarchy, each dimension is associated with some cognitive effort cost $\kappa_{r(i)}$.

The exact loss function is, again, a modeling choice. It should include losses from an imperfect perception and a loss from exerting cognitive effort. Let us consider the following maximization problem:

$$\max_{m_i} (-L) = -\frac{1}{2} \sum_i (1 - m_i)^2 \mu_i - \sum_i \kappa_{r(i)} |m_i|, \quad (14)$$

where μ_i could have several forms depending on the modeling choice. Consider, for example, $x_i^d = \bar{x}_i$ and losses from errors that are quadratic in each dimension. Then one might use $\mu_i = v_i^2 \sigma_i^2$, where σ_i^2 is the variance of attribute i in the set of alternatives. Consider the case $x_i^d = \min_{a \in A} x_i^a$. Then μ_i is the product of v_i and the sum (or the average) increase in value by considering the true value of an alternative. Whatever the modeling choice, as x_i^d is independent of m_i , μ_i is a term capturing a product of v_i and some measure of dispersion of attribute x_i in the set of alternatives.

behavioral impact.

²²One may argue that the DM's objective is to make the right decision, not to form a correct representation of the world. So, one might want to insert the loss from taking the wrong action into the objective function. Yet, to determine that loss one needs to know the right action. The problem would amount to choosing the optimal attention with hindsight. Without the information about the correct action, the best thing one can do is to optimize the representation of the world one bases decisions on. An accurate representation of an alternative's desirability is vital to make the correct choice.

The term capturing cognitive effort cost in a dimension i is the ℓ_1 norm of the attention parameter m_i times the cognitive cost parameter associated with the rank of i .²³

The optimal solution of problem (14) is then given by

$$m_i = \max \{0, 1 - \kappa_i / \mu_i\}. \quad (15)$$

Hence, whenever $\mu_i < \kappa_{r(i)}$, dimension i is neglected. One could thus interpret μ_i as a measure of importance of dimension i to the DM. If the dimension is important enough compared to the cognitive costs associated with its consideration, it will be taken into account. And, given that it is taken into account ($m_i > 0$), the extent to which a dimension is taken into account rises in the importance of the dimension.

It is important to note that the attention m_i which a dimension i receives is crucially determined by the cognitive costs $\kappa_{r(i)}$ associated with its consideration. As was discussed before these costs shall reflect the rising difficulty of solving increasingly complex problems. We will thus assume

$$\begin{aligned} \kappa_r &= 0 && \text{for } r = 1, \\ \kappa_{r+1} &> \kappa_r && \forall r \in \{1, \dots, n-1\}. \end{aligned} \quad (16)$$

The attribute which is considered first, i.e. the attribute which receives rank 1 in the attention hierarchy, is considered without cognitive effort. Considering additional attributes becomes increasingly costly. The attention hierarchy r is thus crucial for the eventual attention allocation. Consider the following 2-step procedure. First, the DM needs to select an attention hierarchy $r : I \rightarrow \{1, \dots, n\}$ which associates each dimension with some consideration costs $\kappa_{r(i)}$. One can think of this as the problem to determine which dimension to consider first, which second, and so on. After assigning a rank to each dimension, the DM solves the above described problem of optimal attention allocation given some assignment of consideration costs. The problem can then be solved backward. Given any assignment $r(i)$, the optimal attention allocation is given by (15).

²³For an elaborate discussion of modeling cognitive costs with such a cost function, see Gabaix (2011).

Plugging this back into the objective function (14) yields

$$\begin{aligned}
(-L) &= \sum_{i:m_i>0} \left[-\frac{1}{2} \left(\frac{\kappa_{r(i)}^2}{\mu_i^2} \right) \mu_i - \kappa_{r(i)} + \frac{\kappa_{r(i)}^2}{\mu_i} \right] - \frac{1}{2} \sum_{j:m_j=0} \mu_j \\
&= \sum_{i:m_i>0} \left[\frac{\kappa_{r(i)}^2}{\mu_i} - \kappa_{r(i)} \right] - \frac{1}{2} \sum_{j:m_j=0} \mu_j.
\end{aligned}$$

The objective at the first stage, anticipating the result of the second stage, is thus

$$\max_{r(i)} \sum_{i:m_i>0} \left[\frac{1}{2} \frac{\kappa_{r(i)}^2}{\mu_i} - \kappa_{r(i)} \right] - \frac{1}{2} \sum_{j:m_j=0} \mu_j. \quad (17)$$

The optimal assignment $r^*(i)$ will satisfy:

$$\mu_i > \mu_j \Rightarrow r(i) < r(j). \quad (18)$$

The optimal assignment will thus assign higher attention ranks to more important dimensions.

Proof. Under an optimal assignment $r^*(i)$ interchanging the ranks of any two attributes $i, j \in I$ may not lead to an increase in $(-L)$. Note that the objective function $(-L)$ is additively-separable across attributes. We may thus confine attention to the parts of the objective function that depend on the two attributes i and j .

Suppose $\mu_i = \mu_j$. It is easy to see that interchanging their ranks has no effect on the objective function. We will thus only look at cases in which $\mu_i > \mu_j$. Take any ranking r . Under this ranking attribute i and j are associated with some cognitive costs $\kappa_{r(i)}, \kappa_{r(j)}$. Denote by $\kappa_h = \max \{ \kappa_{r(i)}, \kappa_{r(j)} \}$ and $\kappa_l = \min \{ \kappa_{r(i)}, \kappa_{r(j)} \}$. Whichever attribute is assigned κ_l has a higher rank under r . We will now show that if r does not assign κ_l to attribute i (the one with strictly higher importance), r cannot be a maximizer for $(-L)$ for some set of cognitive costs $(\kappa_1, \kappa_2, \dots)$ satisfying our assumption (16).

Let us distinguish four cases:

- (i) $\kappa_h > \kappa_l > \mu_i > \mu_j$. Both attributes are neglected before and after interchanging the rank.

The objective function is thus invariant to such a change in ranking.

(ii) $\mu_i > \mu_j > \kappa_h > \kappa_l$. Both attributes are taken into account at the lower rank. However, it is better to assign attribute i the higher rank (and thus κ_l) if

$$\begin{aligned}
& \frac{1}{2} \frac{\kappa_l^2}{\mu_i} - \kappa_l + \frac{1}{2} \frac{\kappa_h^2}{\mu_j} - \kappa_h > \frac{1}{2} \frac{\kappa_l^2}{\mu_j} - \kappa_l + \frac{1}{2} \frac{\kappa_h^2}{\mu_i} - \kappa_h \\
\Leftrightarrow & \frac{\kappa_l^2}{\mu_i} + \frac{\kappa_h^2}{\mu_j} > \frac{\kappa_l^2}{\mu_j} + \frac{\kappa_h^2}{\mu_i} \\
\Leftrightarrow & (\kappa_h^2 - \kappa_l^2)\mu_i > (\kappa_h^2 - \kappa_l^2)\mu_j \\
\Leftrightarrow & \mu_i > \mu_j.
\end{aligned}$$

(iii) $\kappa_h > \mu_i > \mu_j > \kappa_l$. Both attributes are considered at the higher rank but neglected at the lower rank. It is better to assign attribute i the higher rank (and thus κ_l) if

$$\begin{aligned}
& \frac{1}{2} \frac{\kappa_l^2}{\mu_i} - \kappa_l - \frac{1}{2} \mu_j > \frac{1}{2} \frac{\kappa_l^2}{\mu_j} - \kappa_l - \frac{1}{2} \mu_i \\
\Leftrightarrow & \kappa_l^2 \mu_j - \mu_j^2 \mu_i > \kappa_l^2 \mu_i - \mu_i^2 \mu_j \\
\Leftrightarrow & \mu_i \mu_j (\mu_i - \mu_j) > \kappa_l^2 (\mu_i - \mu_j) \\
\Leftrightarrow & \mu_i \mu_j > \kappa_l^2, \text{ which is true since } \mu_i > \mu_j > \kappa_l.
\end{aligned}$$

(iv) $\mu_i > \kappa_h > \mu_j > \kappa_l$. Attribute i is considered both at the higher and lower rank. Attribute j is only considered at the higher rank but neglected at the lower rank. Still, it is better to assign attribute i to the higher rank if

$$\frac{1}{2} \frac{\kappa_l^2}{\mu_i} - \kappa_l - \frac{1}{2} \mu_j > \frac{1}{2} \frac{\kappa_l^2}{\mu_j} - \kappa_l + \frac{1}{2} \frac{\kappa_h^2}{\mu_i} - \kappa_h.$$

This can shown to hold, since

$$\begin{aligned}
& \frac{1}{2}\kappa_l^2 \left(\frac{1}{\mu_i} - \frac{1}{\mu_j} \right) - \frac{1}{2}\mu_j - \frac{1}{2}\kappa_h^2 \frac{1}{\mu_i} + \kappa_h = \\
& \frac{1}{\mu_i\mu_j} \left[\frac{1}{2}\kappa_l^2(\mu_j - \mu_i) - \frac{1}{2}\mu_i\mu_j^2 - \frac{1}{2}\kappa_h^2\mu_j + \kappa_h\mu_i\mu_j \right] = \\
& \frac{1}{2\mu_i\mu_j} \left[\kappa_l^2(\mu_j - \mu_i) + (\kappa_h\mu_i\mu_j - \mu_i\mu_j^2) + (\kappa_h\mu_i\mu_j - \kappa_h^2\mu_j) \right] = \\
& \frac{1}{2\mu_i\mu_j} \left[\kappa_l^2(\mu_j - \mu_i) + \mu_i\mu_j(\kappa_h - \mu_j) + \kappa_h\mu_j(\mu_i - \kappa_h) \right] > \\
& \frac{1}{2\mu_i\mu_j} \left[\kappa_l^2(\mu_j - \mu_i) + \kappa_h\mu_j(\kappa_h - \mu_j) + \kappa_h\mu_j(\mu_i - \kappa_h) \right] = \\
& \frac{1}{2\mu_i\mu_j} \left[(\mu_j\kappa_h - \kappa_l^2)(\mu_i - \mu_j) \right] > 0,
\end{aligned}$$

in which the first inequality (line 4) holds since $\mu_i > \kappa_h > \mu_j$, so the middle term is replaced by strictly lower term. The final inequality holds since $\mu_i > \kappa_h > \mu_j > \kappa_l$.

We have now shown that an optimal solution is to set $r(i)$ according to (18). One may argue that this rule may not be obeyed for dimensions for which case (i) holds. While this is true, one can counter that (18) is optimal for a choice problem (A, v) , and thus for a given vector of $(\mu_i)_{i=1..n}$, for *all* cognitive cost vectors $(\kappa_1, \dots, \kappa_n)$ for which $\kappa_{r+1} > \kappa_r, \forall r = 1..n$. The optimal assignment (18) is thus invariant to changes in the cost vector (e.g. due to changes in cognitive resources for some given choice task). In addition, even if case (i) may hold for some attributes for a given cost vector, it cannot hold for all attributes as long as $\kappa_1 = 0$. \square

8.1 Characteristics of the Attention Function

This section seeks to discuss some characteristics of the attention function, that we derived.

First, more important attributes receive (weakly) more attention than less important attributes, $\partial m_i / \partial \mu_i \geq 0$. The attention each attribute receives thus depends positively on its value to the decision-maker. In addition to this internal factor, the attributes dispersion within the choice set A increases attention. The attention an attribute receives thus depends on the choice environment. It is thus possible to **attract attention** to an attribute by varying the choice set appropriately.

Second, as one attribute gains importance it may eventually gain rank in the attention hierarchy. As another attribute receives a lower rank this other attribute loses attention, $\partial m_i / \partial \mu_j \leq 0$, with strict inequality only if $\partial m_j / \partial \mu_j > 0$. It is hence possible to **distract attention** from an attribute. It is noteworthy that this distraction effect only works through the attention hierarchy and is thus discontinuous. While this feature might be mathematically undesirable, it has some desirable effects which we will discuss shortly.

Next, the attention process features **neglect**, or, in Gabaix' terminology, the attention vector is sparse. Technically, for any decision problem (A, v) there exist vectors of cognitive costs κ satisfying our assumptions such that there exist attributes $i \in I : m_i = 0$ whenever $|I| \geq 2$. So, for any complex choice problem, that is one which involves at least two dimensions, cognitive costs may lead to the neglect of at least one dimension. Similarly, for any vector of cognitive costs κ satisfying our assumptions there exist choice problems $(A, v) : |I| \geq 2$, such that at least one of the dimensions is ignored.

In addition, due to our assumptions on κ , for any choice problem (A, v) there exist attributes $i \in I : m_i > 0$. So, there is **no complete neglect**. As we seek to model the need to simplify a complex choice problem, the DM always considers at least one dimension as this amounts to solving a simple choice problem. This directly implies that complexity costs, as modeled here, will never lead to strictly dominated choices.

The attention hierarchy is not just implicit. Any two attributes that are considered receive a **different weight**: $m_i \neq m_j \forall i, j \in I : m_i, m_j > 0$. More specifically, for any two attributes that are considered, the higher ranking attribute receives strictly more attention.

This, together with the impossibility of complete neglect, implies that the attention process *always* features **over- and underweighting**. Let $\bar{m} = \frac{1}{n} \sum_i m_i$. Then for any complex choice problem and cognitive cost vector κ satisfying our assumptions there exist attributes which are overweighted and attributes which are underweighted. Formally, $\forall (A, v, \kappa) : \exists i \in I : m_i > \bar{m}$ and $\exists j \in I : m_j < \bar{m}$. This is important as it implies that under the derived attention function the decision utility of an alternative is not just an affine transformation of hedonic utility.

8.2 Final Remarks

The optimal attention allocation rule

$$m_i = \max \left\{ 0, 1 - \frac{\kappa_r(i)}{\mu_i} \right\}$$
$$\mu_i > \mu_j \Rightarrow \kappa_r(i) < \kappa_r(j) \quad \forall i, j \in I$$

we derived is, of course, more general than the attention allocation rule we employ in the main part of this paper. We have decided to opt for $\mu_i = \max_{a \in A} - \min_{a \in A}$ as a simple measure of dispersion as it yields unique results for the optimal product design. The results will remain qualitatively unchanged when employing a different measure of dispersion. In addition, we have assumed a tie-breaking rule $\mu_i = \mu_j$ and $v_i > v_j \Rightarrow \kappa_r(i) < \kappa_r(j) \quad \forall i, j \in I$ for the determination of the attention hierarchy. Assuming a different tie-breaking rule would not change our results qualitatively.