Persuasion Through Ordered Information

Sergey Turlo*†

July 20, 2019

Abstract

I consider a model of costly information acquisition where the order in which a receiver can learn about the various characteristics of a multi-attribute offering is determined by a sender who wants to persuade the receiver to accept the offering. I characterize optimal attribute orders when the receiver either learns simultaneously or sequentially, and attributes are heterogeneous either with respect to learning costs or dispersions. Further, I obtain clear-cut conditions on when there is an over- and under-provision of information and show that the number of attributes learned in equilibrium is non-monotonic in the receiver's outside option. Moreover, this paper highlights the role of the distribution of costs and dispersions across attributes as a key determinant for the sender's ability to persuade. Lastly, I discuss various applications such as the choice of obfuscation and the regulation of information provision.

Key words: persuasion, costly information processing, ordered information acquisition, multi-attribute offers

JEL: D83, D18, M31

*Johann Wolfgang Goethe University Frankfurt. E-mail: sergey.turlo@hof.uni-frankfurt.de.

†I thank my advisor Roman Inderst for valuable comments and continuous guidance. The paper has benefited from discussions with Ester Faia, Michael Grubb, Florian Hoffmann, Michael Kosfeld, Jan Landwehr, Thomas Otter, Sebastian Schuler, Roland Strausz, and Florian Zimmermann. Further, I thank participants of the management and microeconomics brown bag seminar at Goethe University. Any errors are mine. You can find the most recent version of the paper here: `https://drive.google.com/open?id=15NJ0c70Q7gmCwwfUsQ7_UR3cJ505Nxst`.

1 Introduction

Consider a sender who wants to persuade a receiver, who is a Bayesian updater, to accept an offer. Does the order in which different pieces of information about the offer can be acquired play a role in persuading the receiver? If yes, what is the sender's optimal sequence of information, and which primitives determine the sender's ability to persuade?

Kamenica and Gentzkow (2011) have sparked a growing literature on persuasion deriving implications for optimal information design. The majority of this literature relies on the receiver being a passive consumer of information without any costs of acquiring it. An abundance of evidence, however, indicates that processing information is costly. When, in addition, not all pieces of information are equal, e.g., as some characteristics of an offer are easier understood than others, then the order in which information can be learned is key for the realized information acquisition. It will have an impact both on what is learned and how much is learned. Undoubtedly, this plays a role in a plethora of economic applications.

First, in marketing and advertising, a potential customer is likely to learn about a firm's product initially in an advertisement and only afterwards she will visit the firm's website for further information. The firm has to determine what to communicate through each channel. Second, a developer of a robo-advisor has to design an algorithm that selects when a particular fact about a financial product must shown to an advisee. Third, in order to persuade potential investors, an issuer of securities has to decide about the sequential structure of his road show presentation. Lastly, a researcher who tries to convince a referee has to find the optimal order to present her research results. Note that in all of the above examples the receivers of information have costs of processing it - be it cognitive costs or simply time.

In this paper, I study a model where a sender (he) tries to persuade a receiver (she) to accept a multi-attribute offering. The sender first determines the order in which the receiver can subsequently costly learn about the attributes. One can think, for example, of a product brochure where the manufacturer determines at which page of the brochure to discuss the different attributes or characteristics of his product. Then, the receiver chooses how much to learn, and once the information is acquired, she decides whether to accept or reject the offering. Throughout, there are no transfers of any form between the parties and the sender cannot distort or conceal any information.[1] Moreover, the sender commits to an order and the receiver observes it ex-ante. This paper shows that the order of information has an impact on receiver learning. I characterize sender-optimal orders when the receiver is either able to commit in advance to a learning path ("simultaneous learning") or when she dynamically decides when to stop learning ("sequential learning"). In the baseline model, attributes are heterogeneous with respect to information processing costs, and an extension studies the case when attributes differ in dispersions. I examine how the equilibrium changes in the primitives of the model, such as the receiver's outside option or the distribution of costs across attributes, and I am particularly interested in the determinants of the sender's ability to persuade.

In the baseline scenario, the sender's offering has a finite number of attributes which are heterogeneous with respect to learning costs. The receiver learns simultaneously, that is, after observing the order of attributes she decides in advance how many attributes to learn. For instance, a consumer can commit ex-ante to reading a number of pages of a product brochure before making her purchasing decision. In this model, learning results in an increase of the dispersion of the posterior expectation in the form of a single-crossing mean-preserving spread, which is also known as a "rotation" in the literature[2]. This property implies that the sender's payoffs are decreasing in the receiver's "learning decision," that is, the number of attributes she commits to learning, when the receiver's outside option is smaller than the prior mean of the offering. Conversely, when the outside option is large as compared to the prior mean, the sender's payoffs are increasing in the learning decision. Thus, the rotation property allows to clearly identify when there will be an over- or an under-provision of information in equilibrium. The sender-optimal learning decision k^* can always be achieved by letting the receiver learn first about the k^* least costly attributes, being ordered from most costly to least costly, followed by the sequence of the (remaining) most costly attributes, also ordered from most to least costly. This sequence of attributes maximizes the receiver's incentives to learn the equilibrium number of attributes and maximizes her expected payoff given the sender's objective to induce a particular learning decision in equilibrium.

[1] Regulation can require the sender to provide certain pieces of information. Alternatively, one can think of a two stage information provision process where the sender first determines what to communicate and then chooses the order in which information can be acquired.

[2] See Johnson and Myatt (2006) for further details.

The relationship between the receiver's outside option and her equilibrium learning decision is shown to be inverse U-shaped. This relates to the fact that as the outside option approaches, either from below or from above, the value of the prior mean the likelihood of making a false decision increases and thus the learning incentives increase. While it is intuitive that the equilibrium learning decision is decreasing in the learning costs, this paper shows that not only the sum of learning costs affects the equilibrium outcome but also their distribution across attributes. In particular, when the most costly attributes become costlier at the expense of the least costly attributes, which can be interpreted as an increase in the concentration of the cost distribution, the set of learning decisions that can be induced by the sender grows and therefore also the sender's ability to persuade improves. As a result, the sender's expected payoff always increases in response to such a shift in the distribution of costs.

Clearly, some situations come to mind where the assumption of committing in advance to learning a specific amount of information appears to be rather restrictive. For instance, while reading a product brochure, a consumer can stop at any time and make her purchase decision based on the information acquired until then. I therefore also study the case of sequential learning where after learning the realization of an attribute the receiver can choose to stop learning. Effectively, the receiver's learning problem becomes one of optimal stopping. While this is clearly interesting, it leads to analytical challenges. In particular, the resulting optimal stopping problem is inherently non-stationary due to the heterogeneity of the attributes which is key for orders to have an impact. Thus, I restrict my analysis to the case where the offering has two attributes. Changing the order of attributes affects both the receiver's incentives to start learning and the likelihood that she continues learning. This, in turn, shows how the sender's expected payoff is affected by the choice of order. Similar to the baseline case, the sender's objective is to prevent learning when the receiver's outside option is relatively small, and to encourage learning when it is rather large. Analogously to above, the expected number of attributes learned in equilibrium is inverse U-shaped in the receiver's outside option, and it declines as learning costs increase. While the choice of the optimal order is still driven by both the sum of costs and their distribution across attributes, it is now harder to characterize how changes in the cost distribution affect equilibrium outcomes. The reason for this is that cost shifts that keep the sum of costs constant no longer keep the value of starting to learn constant.

I apply this form of persuasion to three examples to illustrate how the analysis of economic applications is enriched by the consideration of information orders. First, I study how a firm that sells its product through sales agents chooses the level of obfuscation for the product's individual attributes. Here, the key assumption is that the sales agent learns the consumer's outside option which is ex-ante, and thus at the time of the obfuscation choice, unknown to the firm. Second, I examine how the regulation of information should take into account the possibility of persuasion through ordered information by firms.[3] Lastly, I embed this (bilateral) model of persuasion into a simple search market which allows to endogenize the receiver's outside option and to study the difference between the impact of search costs for new offerings and learning costs on equilibrium outcomes.

A key assumption of the paper is that both the sender and the commit to an order. In the case of the sender this appears to be a natural assumption in many settings as indicated by the introductory paragraphs. In the case of the receiver this assumption appears to be less innocuous. One can argue that receivers are able to optimally pick the most relevant pieces of information in the sense of rational inattention. In certain cases, e.g., during a presentation, it is still the sender who fully controls the order of information, even when the audience is rationally inattentive. Second, even in more interactive settings, it is likely that in particular less sophisticated or less experienced receivers prefer to stick to the order as chosen by the sender which has important implications for consumer protection.[4]

Lastly, I briefly discuss how the results that are derived based on the assumption of attribute heterogeneity with respect to learning costs extend to the case when attributes differ in their dispersions.

Literature This paper relates to several strands of literature. First, it contributes to the literature on (match-specific) information provision by senders, such as Johnson and Myatt (2006), Ganuza and Penalva (2010), and Kamenica and Gentzkow (2011), where a sender decides how much information to provide to a receiver about a state that is uncertain to both parties. In modeling the learning process, these papers follow either a

[3]Notably, firms can steer consumer learning by making some pieces of information more salient. This issue has gained attention by regulatory authorities as well. For instance, in December 2016 a new EU regulation (No 1169/2011) on the provision of food information to consumers has started to apply. It addresses, among other things, the legibility of information and the harmonization of information presentation for certain pieces of information.

[4]Note, however, that there is ample evidence that even professional receivers, such as doctors, accountants, or institutional investors, can be affected by the order in which information is presented.

reduced form approach, impose a rather simplistic learning environment, or allow for an unconstrained choice of the information structure. The current paper contributes to this literature by studying a non-trivial microfoundation of that process. This, in turn, allows me to address the relationship between the properties and features of an offer and the informational outcome in a strategic environment.

Second, this paper is linked to a recent literature that studies ordered acquisition of information. One branch of this literature deals with ordered consumer search.[5] Related to the present paper is the work by Gamp (2017) that studies a multi-product monopolist who determines through obfuscation the search costs for the respective products and thus affects the order in which a buyer searches. Note that in this part of the literature the consumer obtains information about different products, whereas in my model the receiver learns more information about a single product. A second smaller strand of this literature deals with ordered learning about products with multiple attributes. My paper is most notably related to Klabjan et al. (2014) who study the optimal order of information acquisition of a single agent facing the decision whether to accept a multi-attribute object. The current work contributes to this literature by analyzing a setting where the order in which information about a single offer can be learned is chosen strategically by a counterparty.

Lastly, there is a large body of empirical evidence showing that the order in which information is presented has an impact on decision making. This evidence originates from various areas such as marketing (Dominique-Ferreira, 2017), financial reporting (Baird and C. Zelin, 2000), accounting (Hellmann et al., 2017), and even medical science (Bergus et al., 1998).[6] Interestingly, several of those studies have shown that even sophisticated and experienced receivers can be affected by the order of information. Moreover, senders are aware of this and, therefore, use the presentation of information strategically to influence receivers.[7]

[5]For instance, Armstrong et al. (2009) consider a model where a single prominent firm is visited by all consumers first, while Arbatskaya (2007) analyzes a market where a consumer has to visit shops in a predetermined order. For a survey see Armstrong (2017).

[6]Most of these studies make use of the seminal work by Hogarth and Einhorn (1992) who provide a descriptive theory of belief updating that is based on anchoring and belief adjustment. Implications of this theory are that task properties, such as task complexity, task length, and the response mode, determine the existence and the type of order effects.

[7]In a related context, Levav et al. (2010) study the impact of attribute orders with product customization with durable goods. They, however, focus on the choice of a *default attribute* rather than the impact on accepting an outside option.

Organization This paper proceeds as follows. I study optimal orders with simultaneous learning in Section 2 and with sequential learning in Section 3. Section 4 applies this model of persuasion to several examples. Section 6 examines how results translate to a setting where attributes differ in their dispersions and Section 5 discusses key assumptions of this paper. Section 7 concludes. All proofs are in Appendix A. Additional material is provided in Appendix B.

2 The Baseline Model: Simultaneous Learning

The baseline model of this paper considers the case where the receiver learns simultaneously, that is, after observing the order in which learning takes place the receiver commits to learning a number of attributes. Here, attributes differ in the costs that are necessary to learn their actual value. This is motivated by cases where certain pieces of information are easier (and faster) understood than others. For example, in the context of a financial advisor, it is easier to understand who issues a pension plan than its legal terms and conditions.

Problem Setup Consider a sender (he) and a receiver (she). The receiver chooses between an offering by the sender and an outside option. The offering has N attributes v_i and provides an uncertain payoff of $u = \sum_{i=1}^{N} v_i$. I start my analysis with the case where the attributes v_i are identically and independently distributed according to a distribution function $F(\cdot)$. Suppose that $F(\cdot)$ has convex support on $[\underline{v}, \overline{v}]$ where I allow for $\underline{v} = -\infty$ and $\overline{v} = \infty$ and the corresponding density $f(\cdot)$ is continuous, unimodal, and symmetric around mean 0. Learning about an attribute v_i comes at costs $s_i \geq 0$ and attributes differ in their learning costs with $0 \leq s_1 < s_2 < ... < s_N$. Both the sender and the receiver are risk neutral and there is no discounting.

The outside option provides the receiver with a fixed payoff of $\underline{v} < R < \overline{v}$ which is commonly known. The sender obtains a payoff of 1 if the offering is chosen by the receiver and 0 otherwise. There are no transfers between the sender and the receiver.[8] The key assumption of this paper is that the order in which the receiver can learn about the attributes is determined by the sender and known to the receiver before she makes her

[8] Note that the model can be relabeled in order to allow for an exogenous transfer between the sender and the receiver.

learning decision, that is, the number of attributes she wishes to learn.

The order (or the sequence) in which the receiver can observe the attributes is denoted by σ where the expression $\sigma(i)$ denotes the index of the attribute that is learned at the i-th position of the order σ. Formally, σ is a one-to-one correspondence from the set $\{1, ..., N\}$ to itself. When necessary, I explicitly spell out the order in vector notation so that $\sigma = (\sigma(1), ..., \sigma(N))$. To clarify the notation consider the following example. Suppose $N = 3$ and the order chosen by the sender is σ with $\sigma(1) = 1$, $\sigma(2) = 3$, and $\sigma(3) = 2$ or in short $\sigma = (1, 3, 2)$. If the receiver facing σ decides to learn about two attributes, then she will learn about attributes v_1 and v_3 but not about attribute v_2. Denote with $\mathcal{S}$ the set of all orders σ.

The timing of the game is as follows. First, the sender chooses the order in which the attributes can be investigated. Then, the receiver observes this order and decides how many attributes to learn (learning decision). Lastly, the receiver decides, based on this information, whether to accept the offering or her outside option.

Note that this is a game of symmetric information. Before any information has been obtained, in particular when the sender picks the order, both parties only know the prior distribution of the attributes but not their respective realization. When the receiver learns, only she updates her beliefs.

Receiver Problem The receiver will accept the offering after learning T attributes only if her posterior valuation $V(T) = \sum_{i=1}^{T} v_{\sigma(i)}$ exceeds her outside option, $V(T) \geq R$. When $T = 0$, her posterior expectation equals her prior expectation: $V(0) = E[u] = 0$. I denote with $H_T(\cdot)$ the cdf of $V(T)$. Observe that the order σ affects the receiver's expected payoff only through costs $s_{\sigma(i)}$. The posterior expectation $V(T)$ and its cdf $H_T(\cdot)$ do not depend on the order σ directly since all attributes share the same distribution. In equilibrium, however, the order will clearly have an impact on the distribution through the receiver's learning decision.

The receiver's learning problem is a problem of static information acquisition. The receiver's objective is to choose the number of attributes T given the order σ that maximizes her expected net payoff

$$\max_{T \in \{0,1,..,N\}} E\left[\max\{R, V(T)\}\right] - \sum_{i=1}^{T} s_{\sigma(i)}. \tag{1}$$

Denote with $MB_i = E[\max\{R, V(i)\}] - E[\max\{R, V(i-1)\}]$ the marginal benefits of learning the i-th attribute gross of costs and let $T^*(\sigma)$ be the solution to the receiver's information acquisition problem. If the receiver is indifferent between two learning decisions T and T', she chooses the learning decision that maximizes the sender's payoff.

Sender Problem The sender is risk-neutral and has no costs of production or adjusting the order of attributes. Thus, the sender's payoff Π equals the probability that the receiver accepts his offering. Formally, the sender's objective is to find an order $\sigma \in \mathcal{S}$ that maximizes

$$\Pi(\sigma) = \Pr\left(V(T^*(\sigma)) \geq R\right), \tag{2}$$

where the sender anticipates the receiver's optimal learning decision $T^*(\sigma)$. In what follows, a learning decision k is called *implementable* if an order σ exists such that given the order σ the receiver optimally learns k attributes: $T^*(\sigma) = k$. Denote with $\mathcal{K}_R$ the set of implementable learning decisions for a given outside option R. I impose that when the sender is indifferent between two orders, he picks the order that maximizes the receiver's expected payoff.

Analysis It is well-known in the literature on economics of information that an increase in information, e.g., in the form of more informative signals, results in a more dispersed distribution of the posterior expectation. In this model, a particular dispersion order arises that allows for a clear-cut analysis.

Definition 1 (Rotation) *Suppose two random variables v_i and v_j are endowed with cdfs $F_i(v)$ and $F_j(v)$, respectively, and suppose $E[v_i] = E[v_j] = \mu$. Then, v_i is obtained from v_j by a rotation around the mean μ if $F_i(v) > F_j(v)$ for $v < \mu$ and $F_i(v) < F_j(v)$ for $v > \mu$.*

When the receiver commits ex-ante to learning an additional attribute, the dispersion of her posterior expectation increases in the sense of a rotation.[9]

Lemma 1 *Suppose the receiver learns simultaneously about the sender's offer. Then, her posterior expectation $V(T+1)$ is obtained from $V(T)$ by a rotation around mean 0.*

[9] The rotation property is a commonly used tool in economics of information, see for example Johnson and Myatt (2006) and Hoffmann and Inderst (2011).

Lemma 1 shows that the sender has a clear objective: When the outside option is small, $R < 0$, he will choose the order such that the receiver learns as little as possible since (additional) information increases the likelihood that the receiver learns something that induces her to reject the offering. When the outside option is large, $R > 0$, the contrary is true, and the sender tries to induce as much learning as possible. In this case, information increases the likelihood that the receiver learns something in favor of the offering.

While, clearly, the relation between the costs of the attributes and the benefits of learning will determine the receiver's learning decision, it is not immediately obvious whether the sender's choice of the order will have an impact on the receiver's decision. Intuitively, the order should have an impact when there is sufficient variation in the learning costs. The next Lemma provides sufficient conditions for the sender's order choice to affect the receiver's learning decision.

Lemma 2 *When the receiver learns simultaneously and costs s_i are such that $s_1 < \min_i\{MB_i\}$ and $s_N > \max_i\{MB_i\}$, there exist orders σ and σ' that result in different learning decisions, that is, $T^*(\sigma) \neq T^*(\sigma')$.*

While Lemma 1 characterizes the sender's objective, it remains silent on how the sender can actually achieve the optimal outcome. The next result, therefore, describes the set of orders that contains an optimal order.

Lemma 3 *If learning k attributes is implementable, that is an order σ exists so that $T^*(\sigma) = k$, then this learning decision can also be implemented with the order*

$$\sigma_k(i) = \begin{cases} k+1-i & \text{for } 1 \leq i \leq k \\ N+k+1-i & \text{for } k+1 \leq i \leq N, \end{cases} \tag{3}$$

where $\sigma_0 \equiv \sigma_N$, so that $T^*(\sigma) = T^*(\sigma_k) = k$.

Figure 1 provides an illustration of the cost structures for orders σ_3 and σ_7 when $N = 10$. The intuition for this result is as follows. If it is optimal for the receiver to learn k attributes given some order σ, then learning k attributes instead of $l < k$ attributes increases the receiver's benefits by $\sum_{i=l+1}^{k} \left[MB_i - s_{\sigma(i)}\right] \geq 0$. With the order σ_k learning k attributes instead of $l < k$ attributes becomes more beneficial since the costs of doing so are lower with σ_k as it holds now $\sum_{i=l+1}^{k} s_{\sigma(i)} \geq \sum_{i=l+1}^{k} s_{\sigma_k(i)} = \sum_{i=1}^{k-(l+1)} s_i$. Similarly, the costs of learning $m > k$ attributes are higher with the order σ_k. Effectively, the order σ_k maximizes the incentives to learn k attributes. Thus, if learning k attributes with the order σ is optimal, then it is also optimal to do so with σ_k.

Note also that this result is completely independent of the shape and the curvature of the marginal value of learning which is significant in light of the fact that the value of information can in general be both convex and concave.[10] Clearly, the sender is indifferent between any two orders that implement the same learning decision. Therefore, Lemma 3 simplifies the sender's problem as he faces now an optimization over N different options rather than $N! = |\mathcal{S}|$. Additionally, when the sender implements in equilibrium a learning decision k^*, then the order σ_{k^*} maximizes the receiver's expected payoff.

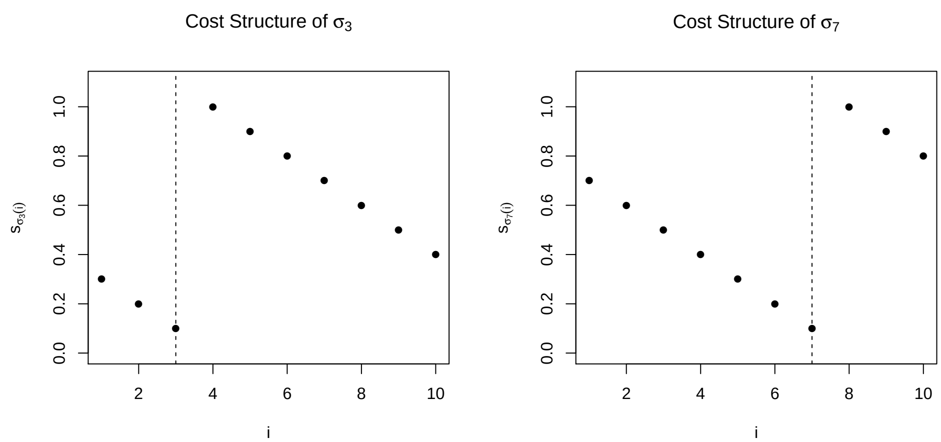

Figure 1: **Cost structures with orders σ_k with $k = 3$ and $k = 7$.** This figure depicts the cost structures of the order σ_3 (left panel) and σ_7 (right panel) when there are $N = 10$ attributes and learning costs are given by $s_i = 0.1i$. The dashed line identifies $k = 3$ and $k = 7$, respectively.

I obtain a benchmark for the equilibrium learning decision by characterizing the order that maximizes the receiver's expected payoff in the next Lemma.

Lemma 4 *When the receiver learns simultaneously about the sender's offer, the attribute order $\sigma^{**}(i) = i$ maximizes the receiver's expected net payoff.*

According to Lemma 4, the receiver's expected payoff is maximized when she can learn in an order where attributes are arranged from least costly to most costly. Recall that the

[10]See for additional information Radner and Stiglitz (1984) and Chade and Schlee (2002).

attribute order has no impact on the gross marginal benefits of learning and so this order minimizes the costs of learning any number of attributes.

Note that the sender can always induce the receiver to learn $k^{**} \equiv T^*(\sigma^{**})$ attributes. Therefore, when the receiver's outside option is small, $R < 0$, the receiver will learn at most k^{**} attributes in equilibrium and so an over-provision of information, relative to the receiver's optimal outcome k^{**}, will never occur. In contrast, when $R > 0$, the opposite is the case and the receiver learns in equilibrium at least k^{**} attributes so that an under-provision will never happen. I can use the results up to now in order to characterize the equilibrium outcome.

Proposition 1 *When the receiver learns simultaneously in the order chosen by the sender, the equilibrium is characterized as follows. The number of attributes learned by the receiver in equilibrium is uniquely determined by* $k^* = \min\{\mathcal{K}_R\}$ *when* $R < 0$*, and it is* $k^* = \max\{\mathcal{K}_R\}$ *when* $R > 0$*, and the sender chooses the order* σ_{k^*} *as defined in (3). There can be an under-provision of information,* $k^* < k^{**}$*, only when the receiver's outside option is small,* $R < 0$*, and there can be an over-provision of information,* $k^* > k^{**}$*, only if her outside option is large,* $R > 0$*.*

Impact of the Outside Option In what follows, I consider how the primitives of the model - the receiver's outside option and the attributes' costs - affect the receiver's equilibrium learning decision. I start with a preliminary analysis where I study how the outside option affects the incentives to learn.

Lemma 5 *When the receiver learns simultaneously, the gross marginal benefits of learning the* i*-th attribute,* MB_i*, increase for all* i *in the outside option* R *when* $R < 0$*, and they decrease in* R *when* $R > 0$*.*

Intuitively, as the outside option R approaches the prior mean, it becomes more likely that the optimal action based on the prior information is incorrect. Therefore the value of information increases. Clearly, the value of learning gross of costs is highest when the outside option equals the prior mean: $R = 0$.

Corollary 1 *The equilibrium number of attributes learned by the receiver,* k^**, is increasing in the outside option* R *when* $R < 0$*, and it is decreasing when* $R > 0$*.*

With Lemma 5, the gross marginal benefits of learning increase as R approaches the prior mean 0 so that for all orders σ the respective optimal learning decisions $T^*(\sigma)$ are increasing in the outside option R when $R < 0$, and they are decreasing when $R > 0$. Therefore, when $R < 0$, also the smallest implementable learning decision $k^* = \min\{\mathcal{K}_R\}$ must be increasing in R, and, similarly, when $R > 0$, the largest implementable learning decision $k^* = \max\{\mathcal{K}_R\}$ must be decreasing in R.

Impact of Costs Next, I consider how the cost structure affects the equilibrium learning outcome k^*. In particular, I am interested in how the sender's ability to persuade depends on the distribution of costs across attributes.

Corollary 2 *The equilibrium learning decision k^* decreases when the learning cost s_i increases by $\Delta s > 0$. When the distribution of costs across attributes changes such that the costs s_i decreases by Δs and the costs s_j increases by Δs with $i < j$, the set of implementable learning decisions $\mathcal{K}_R$ grows, and therefore the equilibrium learning decision k^* decreases when $R < 0$ and increases when $R > 0$.*

The intuition for the first part of Corollary 2 is immediate. When costs of learning increase, the incentives to learn decrease, and thus it becomes easier to prevent learning (when $R < 0$) and harder to encourage learning (when $R > 0$). The second part of the Corollary is more intricate. Recall that with $i < j$ it holds before the shift in the cost structure that $s_i < s_j$. Therefore, a change in the distribution of costs where s_i decreases and s_j increases can be seen as an increase in the concentration of the distribution of costs.[11] With Lemma 3, the sender-optimal order is contained in the set of orders $\{\sigma_k\}_{k=0,\dots,N}$. When the costs of information acquisition become more concentrated, this implies that the incentives to learn until k given an order σ_k increase whereas the incentives to learn more than k decrease. Therefore, if a learning decision k was implementable before the shift, it remains implementable afterwards. Moreover, if the increase in concentration is sufficiently strong, the former effect results in a larger learning decision k^* becoming implementable, while the latter effect facilitates the implementation of learning fewer attributes, i.e., a smaller k^*. As a consequence of such a cost shift the sender's ability to persuade increases, that is, it always becomes more likely in equilibrium that the receiver accepts the offering.

[11]Note that Corollary 2 implies that any shift in the cost structure where the most costly attributes become costlier and the least costly ones become less costly - so that in total the sum of costs remains constant - results in additional implementable learning decisions.

Illustration I close the analysis of the baseline model by providing an illustration for the case when there are $N = 2$ attributes that are normally distributed with $E[v_i] = 0$ and $\text{Var}(v_i) = 0.5$. In this case, we have that the gross marginal benefits of learning the first and second attribute are $MB_1 \approx 0.20$ and $MB_2 \approx 0.15$, respectively. Figure 2 shows how the optimal order depends on s_1 and s_2. When the outside option is small, $R < 0$, the sender tries to prevent learning. It is optimal for him to choose the order $\sigma^* = (2, 1)$ only if the sum of costs is larger than the sum of benefits of learning, $MB_1 + MB_2$, and MB_1 is smaller than s_2 as then this order results in no receiver learning. Otherwise, he picks the order $\sigma^* = (1, 2)$ and thus reduces the number of learned attributes by placing the more costly attribute second so that the receiver learns only one (instead of two) attributes.

When the outside option is rather large, $R > 0$, the sender tries to encourage receiver learning. If costs are sufficiently low, he will place the less costly attribute second in order to ensure that both attributes are learned and so the order $\sigma^* = (2, 1)$ is optimal. However, when the sum of costs is sufficiently high, the best the sender can do is to make sure that the receiver learns at least one attribute which he achieves by presenting the less costly attribute first: $\sigma^* = (1, 2)$.

3 Sequential Learning

In this section, I consider the case when the receiver learns sequentially about the attributes of the offering. After learning an attribute, she has to decide whether she wants to accept her outside option, continue learning, or accept the offering. This is clearly an interesting variation of the baseline model since there are situations where the receiver cannot commit ex-ante to acquiring a fixed amount of information. For example, a consumer reading a product brochure can stop doing so at any time. Unfortunately, the resulting learning problem yields a stopping problem that is inherently non-stationary and makes a general analysis intractable. For this reason, I fix in this section the number of attributes to $N = 2$ in order to obtain a tractable problem. Most of the results obtained in the previous Section translate to sequential learning, albeit the underlying mechanism is different due to the dynamic nature of this version of the model. The timing remains as before: The sender chooses an order, the receiver observes this order, and makes her learning decision. Notably, the sender cannot interfere with the receiver learning once she has started to learn. Then, the receiver decides whether to choose the offering or the outside option.

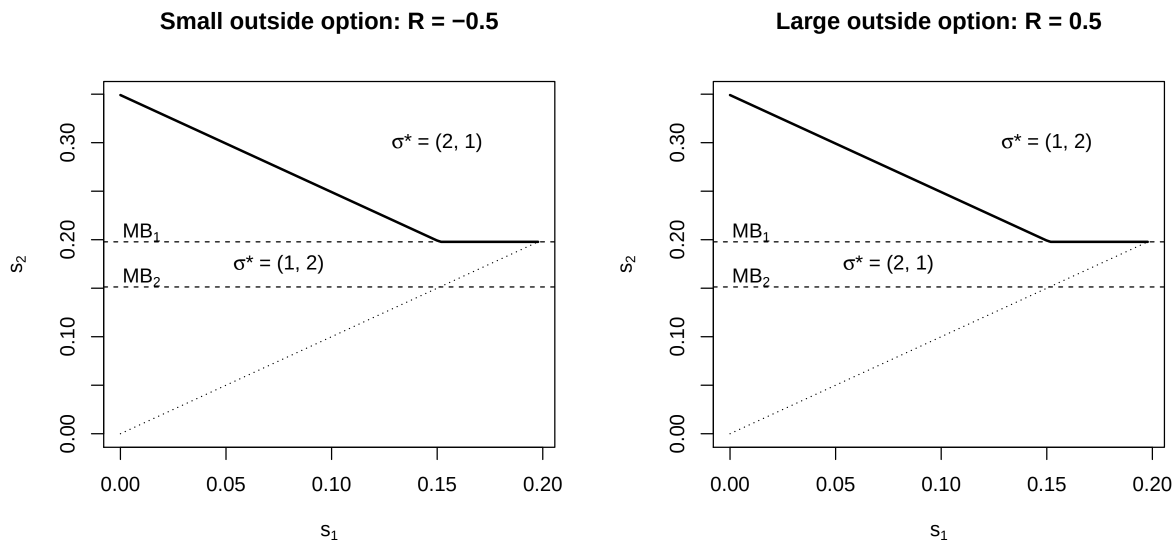

Figure 2: **Illustration of optimal orders with simultaneous learning.** This figure shows the optimal order conditional on the outside option and costs s_1 and s_2 when the sender's offer has two attributes. The outside option is small, $R = -0.5$, in the left panel, and it is large, $R = 0.5$, in the right one. The horizontal dashed lines represent the gross marginal benefits of learning the first and the second attribute, and the dotted lines indicate the points with $s_1 = s_2$. The thick black line separates the parameter space into subsets where $\sigma^* = (1, 2)$ and $\sigma^* = (2, 1)$ are optimal.

Receiver Optimal Stopping The receiver faces an optimal stopping problem given the order in which she can learn about the attributes. Note that now the set of all possible orders is $\mathcal{S} = \{(1, 2); (2, 1)\}$. Denote with H_t the set of histories h_t that are observable by the receiver at the beginning of period t but before the attribute $v_{\sigma(t)}$ has been observed. Define with $\mathcal{H} = \bigcup_t H_t$ the set of all histories.

The receiver's problem is to decide, based on the observable information, when to stop, and, once she has stopped, whether to accept the sender's offering or the outside option. Formally, the receiver has to find a stopping rule $\tau \in \{0, 1, 2\}$ that determines for each history $h_t \in \mathcal{H}$ whether she stops and, if she does so, whether she accepts the offering or the outside option.

Denote with $U_t = E[u|h_t]$ the receiver's expected valuation conditional on the information available in h_t. Clearly, if the receiver stops in t, she chooses the offering if $U_t \geq R$. Thus, the receiver's expected surplus for a stopping rule τ is given by

$$\mathcal{U}(\tau,\sigma) = E\left[\max\{R, U_\tau\} - \sum_{t=0}^{\tau} s_{\sigma(t)}\right] \tag{4}$$

where $s_{\sigma(0)} \equiv 0$ so that there are no costs if the receiver chooses to stop immediately. Therefore, the receiver's problem is to find a stopping rule $\tau^*(\sigma)$ that solves $\sup_\tau \mathcal{U}(\tau,\sigma)$.

Sender's Problem The sender's payoff Π equals the probability that the receiver will accept his offering. Formally, for a given attribute order σ the sender's expected payoff is

$$\Pi(\sigma) = E[\mathbf{1}_{U_{\tau^*(\sigma)} \geq R}] \tag{5}$$

where $\mathbf{1}_{U_\tau \geq R}$ is an indicator that equals one when the receiver accepts the offering and zero otherwise. Note that the sender anticipates the receiver's optimal stopping behavior as indicated by $\tau^*(\sigma)$. As before, the sender has to find an attribute order that maximizes his payoff while taking into account the receiver's optimal stopping. Thus, the sender's problem is given by $\sigma^* = \arg\max_{\sigma \in \mathcal{S}} E[\mathbf{1}_{U_{\tau^*(\sigma)} \geq R}]$.

Analysis In what follows, I first solve the receiver's problem, and then I use this result to study the sender's problem of finding the optimal order of information.

Take an order σ as given. Once the receiver has learned the value of both attributes, she can either accept the offering or take the outside option. She will accept the offering when $v_1 + v_2 \geq R$ and take the outside option otherwise. After she has observed the realization v of $v_{\sigma(1)}$, she has to decide whether to accept the offering, the outside option, or to learn about $v_{\sigma(2)}$ at costs $s_{\sigma(2)}$. Accepting the offering immediately results in a payoff of v, and the outside option provides a payoff R. Finally, the expected value of learning about $v_{\sigma(2)}$ given that $v_{\sigma(1)} = v$ is

$$L_2(v) = E\left[\max\{R, v + v_{\sigma(2)}\}\right] - s_{\sigma(2)}.$$

Intuitively, if $v_{\sigma(1)} = v$ is sufficiently low relative to her outside option R, the receiver takes the outside option R immediately. If it is sufficiently high, she will accept the offering without further learning as this saves her learning costs, and if the valuation v is close to R, the receiver will prefer to learn the value of the remaining attribute $v_{\sigma(2)}$.

Lemma 6 *Consider a receiver who learns sequentially about attributes of the offering in a given order* σ*, and suppose she has already learned about* $v_{\sigma(1)}$*. Then, there are cutoffs* $\underline{b} \leq R \leq \overline{b}$ *such that the receiver accepts the outside option when* $v_{\sigma(1)} < \underline{b}$*, learns about the attribute* $v_{\sigma(2)}$ *when* $\underline{b} \leq v_{\sigma(1)} < \overline{b}$*, and accepts the offering without learning about* $v_{\sigma(2)}$ *when* $\overline{b} \leq v_{\sigma(1)}$*.*

Note that the cutoffs $\underline{b}$ and $\overline{b}$ depend solely on the second attribute $v_{\sigma(2)}$ and its information cost $s_{\sigma(2)}$ and not on the first attribute $v_{\sigma(1)}$. I can use this result to characterize when the receiver starts learning given her prior information. The net value of learning before the first attribute is revealed equals

$$L_1 = F(\underline{b})R + \int_{\underline{b}}^{\overline{b}} L_2(v)dF(v) + \int_{\overline{b}}^{\overline{v}} vdF(v) - s_{\sigma(1)}. \tag{6}$$

As above, the outside option is valued at R whereas the value of accepting the offering is $E[v_1 + v_2] = 0$. Intuitively, the receiver starts learning when costs are sufficiently low.

Lemma 7 *When the receiver learns sequentially, she starts learning only if* $s_{\sigma(1)}$*, the learning cost of the first attribute in the order* σ*, is sufficiently low. Otherwise, when* $R \leq 0$*, the receiver accepts the offering, and when* $R > 0$ *she chooses her outside option without learning.*

Sender Optimal Order As in the baseline, the sender's objective is to maximize the likelihood that the receiver accepts the offering. As there are no further costs on the side of the sender, he does not care how much information the receiver has obtained and what costs she has incurred as long as she accepts offer.

In a preliminary step, I investigate how the sender's payoff changes in $s_{\sigma(2)}$, the learning costs of the second attribute, given that the learning costs of the first attribute $s_{\sigma(1)}$ are so low that the receiver always chooses to observe the first attribute $v_{\sigma(1)}$. Clearly, the cost $s_{\sigma(2)}$ affects the cutoffs $\underline{b}$ and $\overline{b}$ which determine the receiver's stopping behavior after observing the first attribute $v_{\sigma(1)}$. Thus, the sender's payoff equals

$$\Pi = 1 - F(\overline{b}) + \int_{\underline{b}}^{\overline{b}} 1 - F(R - v)dF(v). \tag{7}$$

The following Lemma describes how learning cost $s_{\sigma(2)}$ affects the sender's payoff.

Lemma 8 *Suppose the receiver learns sequentially, and the cost of learning the first attribute, $s_{\sigma(1)}$, is so low that the receiver always learns the first attribute. When $R < 0$ the sender's payoff Π is increasing in the receiver's learning cost of the second attribute, $s_{\sigma(2)}$, whereas when $R > 0$ the sender's payoff is decreasing in $s_{\sigma(2)}$. The relationship is strict when $s_{\sigma(2)}$ is not too high.*

The intuition for Lemma 8 is straightforward. When the learning cost $s_{\sigma(2)}$ increases, the incentive to learn decreases and so $\underline{b}$ increases while $\overline{b}$ decreases. When the outside option is small, $R < 0$, this together with the unimodality as well as the symmetry of the distribution $f(\cdot)$ implies that the increase in the likelihood to accept outweighs the increase in the likelihood to reject after the first attribute is learned. Therefore, the sender's payoff increases. In contrast, when the outside option is relatively valuable, $R > 0$, the sender wants to encourage learning as in this case the effects are reversed and thus his payoff decreases in $s_{\sigma(2)}$.

I can now use the findings up to this point and derive the sender optimal order of attributes. Suppose first that $R < 0$ so that without additional information the receiver prefers the offering to the outside option and recall that $s_1 < s_2$. If the information costs s_2 are sufficiently high to prevent any acquisition of information, then the sender prefers that the receiver learns first the attribute v_2 and then the attribute v_1. If, however, the costs are not sufficiently high, the sender will prefer the receiver to learn first v_1 and only then v_2 as this, following the logic of Lemma 8, increases his payoff. When $R > 0$, the arguments are reversed. In this case, the sender wants to facilitate learning. Thus, if the learning costs are not too high, so that the receiver prefers to start learning, the sender will present first the more costly attribute and then the less costly one: $\sigma^* = (2, 1)$. If, however, costs are too high, the sender prefers to present the attributes in the order $\sigma^* = (1, 2)$ to induce the receiver to learn at least the first attribute so that she accepts the offering with a positive probability. The following Proposition summarizes these results.

Proposition 2 *Suppose the receiver learns sequentially, and the sender determines the order in which learning takes place. Then, the following characterizes the equilibrium outcome. When $R < 0$, there is a decreasing function $a^-(s_1)$ such that when $s_2 > a^-(s_1)$ the sender prefers the order $\sigma^* = (2, 1)$, and the receiver does not learn. When $s_2 < a^-(s_1)$, the sender prefers $\sigma^* = (1, 2)$, and the receiver learns at least one attribute. When $R > 0$,* *there is a decreasing function* $a^+(s_1)$ *such that when* $s_2 < a^+(s_1)$ *the sender chooses* $\sigma^* = (2,1)$*, and the receiver learns at least one attribute. When* $s_2 > a^+(s_1)$*, the sender prefers* $\sigma^* = (1,2)$*, and the receiver learns at least one attribute if learning costs* s_1 *and* s_2 *are sufficiently small.*

Comparative Results Next, I consider the impact of the outside option and the cost structure on the equilibrium outcome. Clearly, the number of attributes learned is stochastic, and so, in the following Corollary, I focus on how the probabilities of learning a number of attributes change. In order to streamline the exposition, suppose for now that there are some values of R such that the receiver learns in expectation more than one attribute regardless of the chosen order.

Corollary 3 *In equilibrium, the probability of learning one attribute and the probability of learning two attributes are both increasing in* R *when* $R < 0$ *and they are decreasing when* $R > 0$*.*

An immediate consequence of this Corollary is that the expected number of attributes learned in equilibrium is inverse U-shaped in the outside option, that is, it is increasing in $R < 0$ and decreasing in $R > 0$. Similar to the case with static learning, as the outside option approaches the prior mean, the value of information increases as the probability to make a wrong (uninformed) decision also increases. This explains the shape of the probability of learning one attribute in R. When the probability of learning two attributes is considered, the intuition is as follows. The probability that the receiver learns the second attribute, conditional on her learning the first one, is given by $F(\bar{b}) - F(\underline{b})$. The cutoffs $\underline{b}$ and $\bar{b}$ increase as the outside option R increases which, in turn, affects the probability of learning the second attribute. With the symmetry and unimodality of the attributes, the probability to learn the second attribute increases in the outside option when $R < 0$, and it decreases otherwise for any order σ. This establishes the inverse U-shaped relation in equilibrium as well. Next, I turn to the impact of costs on the equilibrium outcome.

Corollary 4 *When learning cost* s_i *increases by* $\Delta s > 0$*, the equilibrium likelihoods of learning both one and two attributes decrease. Suppose next that the sum of costs is fixed at* s *so that* $s_1 = s - s_2$*. When the outside option is small,* $R < 0$*, the sender's payoff increases when* s_2 *increases. When the outside option is large,* $R > 0$*, there is a cutoff* $\bar{s}$

such that the sender's payoff increases in s_2 when $s < \overline{s}$. When $s > \overline{s}$, the sender's payoff first increases and then decreases in s_2.

Surprisingly, there is an asymmetry, conditional on the outside option, in how a shift in the distribution of costs affects the sender's ability to persuade. In the case where the outside option is small, an increase in concentration is always beneficial. However, when the outside option is large, $R > 0$, an increase in the concentration of costs is beneficial for the sender only when the sum of costs is sufficiently small. When the sum of costs s is sufficiently large, the impact of concentration is non-monotonic in equilibrium.

When the outside option is small, the sender wants to prevent learning. An increase in the concentration, that is a higher s_2 and a lower s_1, benefits the sender as he can either decrease the incentives to learn the second attribute, with the order $\sigma = (1, 2)$, or reduce the incentives to start learning with $\sigma = (2, 1)$. When the outside option is large, the sender tries to facilitate learning. An increase in the concentration is only beneficial for the sender as long as the receiver starts learning given the order $\sigma = (2, 1)$. Clearly, this will not hold indefinitely, that is, until $s_2 = s$, if the sum of costs s is too large, and therefore, the sender chooses eventually the order $\sigma = (1, 2)$. Then, a further increase in s_2 decreases the receiver's incentives to learn the second attribute.

Illustration Next, I present the results of a simulation where I numerically determine the optimal order for the case when attributes v_1 and v_2 are normally distributed with $E[v_i] = 0$ and $\text{Var}(v_i) = 1$. When the outside option is small, I set $R = -0.25$ and when it is large I set $R = 0.25$.

Figure 3 illustrates optimal orders conditional on the value of the outside option and the learning costs s_1 and s_2. While the simulation clearly confirms the predictions of Proposition 2, it also highlights the role which is played by the concentration of costs. Apparently, the boundaries $a^-(s_1)$ (left panel) and $a^+(s_1)$ (right panel) are strictly decreasing and convex in s_1, and so there are cases when the concentration of costs affects the choice of the optimal order - even though the sum of costs remains constant. An example for this are the points $(s_1, s_2) = (0.15, 0.4)$ and $(s_1, s_2) = (0.25, 0.30)$ in both panels.

Differences between Sequential and Simultaneous Learning I close this section by briefly spelling out the differences between the predictions made by the models with

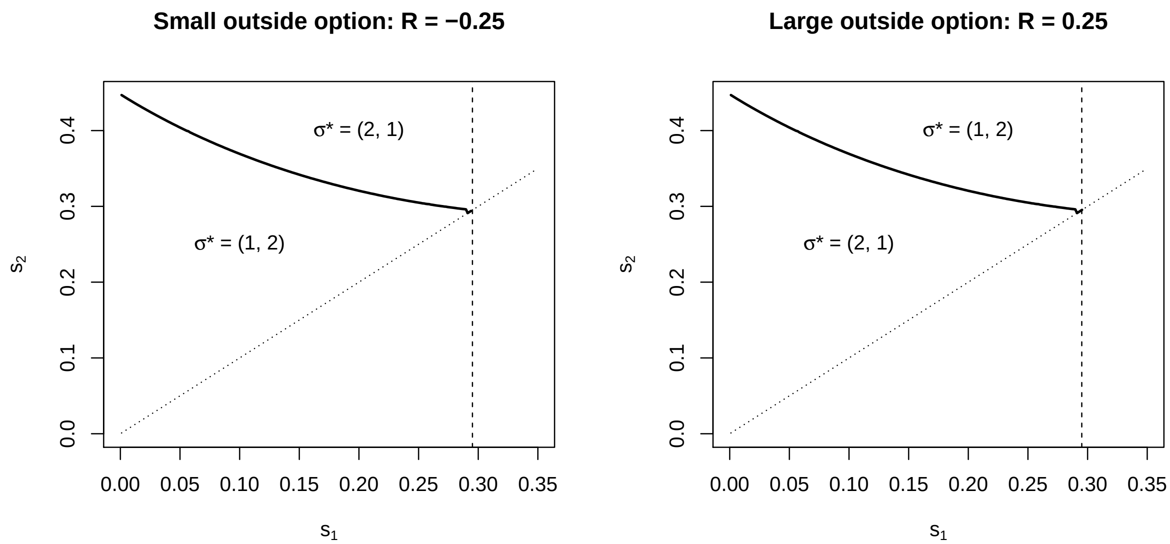

Figure 3: **Illustration of optimal orders with sequential learning.** This figure depicts the optimal order conditional on the outside option and the costs of learning. The left panel depicts the optimal orders when the outside option is small, $R = -0.25$, and the left when it is large, $R = 0.25$. The dotted lines in both panels indicate the set of points with $s_1 = s_2$. The vertical dashed lines indicate the level of costs s_1 such that the receiver never learns. The thick black lines represent $a^-(s_1)$ and $a^+(s_1)$, respectively.

simultaneous and sequential learning. With simultaneous learning, changes in the primitives of the model will always result in discontinuous changes of the outcomes, notably, the expected number of learned attributes and the acceptance probability. With sequential learning, in contrast, there are ranges of the primitives where outcomes change strictly and smoothly in (marginal) changes of the primitives. The reason for this difference is that with sequential learning the learning path of the receiver is not fixed and so she can react to changes in the primitives conditional on what she learns about the first attribute. With simultaneous learning, she lacks this option.

Figure 4 provides an illustration for this difference. It depicts the equilibrium acceptance probability Π^* as a function of the costs of the costlier attribute s_2. In particular, the outside option is $R = -1$ and the offering has two standard normal attributes, $v_i \sim \mathcal{N}(0, 1)$, with $s_1 = 0$. When learning is simultaneous, the equilibrium likelihood will be either constant or have jumps in the costs s_2 such that the resulting shape will resemble a step function. When, in contrast, learning is sequential, the likelihood will have segments where it is continuously increasing in the learning costs s_2. Note that over this range the sender optimally chooses the order $\sigma^* = (1, 2)$ and so a gradual increase of the cost s_2 gradually affects the learning cutoffs $\underline{b}$ and $\overline{b}$. This is then reflected in a gradual increase of Π^* in s_2.

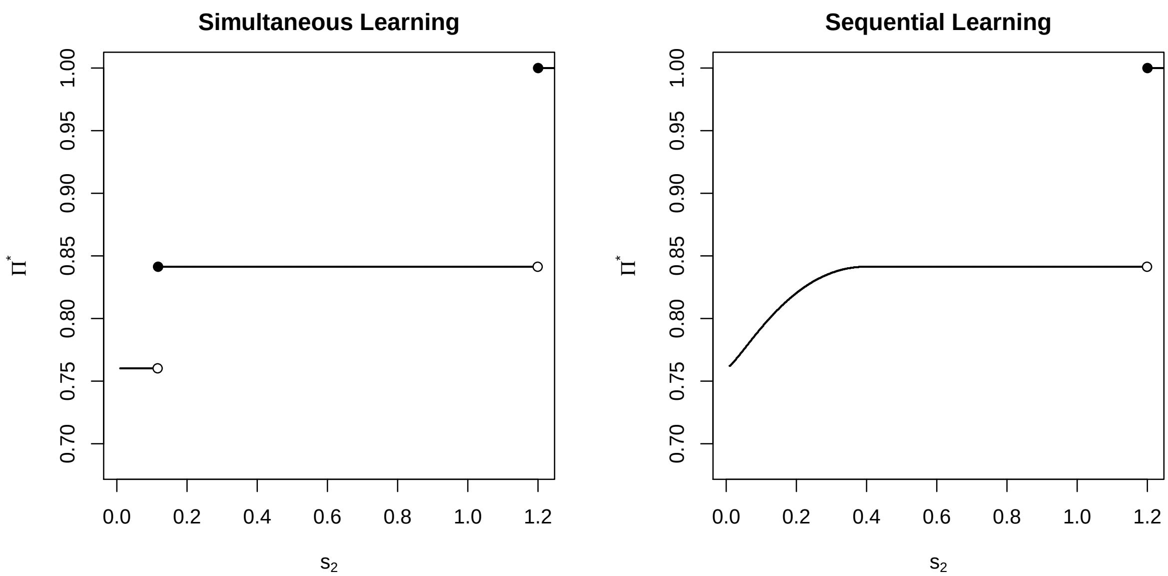

Figure 4: **Acceptance probability with simultaneous and sequential learning in cost s_2.** This figure contrasts how the equilibrium probability of accepting the offering, Π^*, changes in the cost of the more costly attribute s_2 when learning is simultaneous (left panel) and sequential (right panel).

4 Applications

4.1 Choice of Obfuscation Levels

The following application illustrates the link between obfuscation choices by firms and persuasion with ordered information. Consider a firm that offers a product and sells it through a sales agent (sender) to a consumer (receiver). The product has two attributes and the firm chooses the level of obfuscation of each attribute. In particular, it determines the costs s_1 and s_2 that the consumer has to spend in order to learn the value of the respective attribute.[12] Suppose that there are two types of consumers who differ in their

[12] In this Section, I follow the modeling approach by Ellison and Wolitzky (2012) and Gamp (2017) where obfuscation is endogenized by allowing firms to determine the (time) cost that is necessary for consumers to learn about a price or the match value of the offering.

respective outside options: $R_L = -0.5$ and $R_H = 0.5$ with the commonly known probability $\Pr(R = R_H) = \gamma$.[13] When the firm chooses the level of obfuscation, it does not know the value of the consumer's outside option. However, when the sales agent is matched with a consumer, the agent learns it and can thus adjust the order of presentation accordingly. If the agent is able to persuade the consumer, both the firm and the sales agent obtain a payoff of one. Otherwise they obtain zero. The consumer obtains her valuation of the offering $v_1 + v_2$ if she accepts it and her outside option otherwise. The consumer learns simultaneously.

To make things more concrete, suppose attributes follow a standard normal distribution, $v_i \sim \mathcal{N}(0, 1)$ and are identically and independently distributed so that for both consumer types the (rounded) gross marginal value of learning the first attribute is $MB_1 = 0.20$ and of the second attribute it is $MB_2 = 0.15$. When $R = R_L$, the consumer accepts the product with probability 1 when she learns nothing, with probability 0.69 when she learns one attribute, and with probability of 0.64 when she learns two attributes. When she has a large outside option, $R = R_H$, she rejects the offer if she learns nothing, she accepts it with probability 0.31 when she learns one attribute, and accepts with probability 0.36 when she learns two attributes. The firm can pick costs to be either high $s_H = 0.22$ or low $s_L = 0.10$. I summarize the probability of accepting given the outside option and the number of attributes learned in Table 1. In what follows, the sales agent presents the

	$T = 0$	$T = 1$	$T = 2$
$R = R_L$	1	0.69	0.64
$R = R_H$	0	0.31	0.36

Table 1: **Acceptance probability for given outside option and learning decision.**

information in an optimal order. The firm can choose between three different strategies: $s_1 = s_2 = s_L$, $s_1 = s_2 = s_H$, and $s_1 = s_L, s_2 = s_H$. Given that the sales agent optimally adjusts the order of attributes, we obtain the following probabilities of accepting given the consumer's outside and the firm's choice of costs. For example, if the consumer's outside option is large, $R = R_H$, and the firm has picked $s_1 = s_L = 0.10$ and $s_2 = s_H = 0.22$, the sales agent will optimally present first the attribute v_2 and then the attribute v_1.

[13] A motivation for this is that some consumers tend to compare more than others and are, therefore, aware of better alternatives.

	$s_1 = s_L, s_2 = s_L$	$s_1 = s_L, s_2 = s_H$	$s_1 = s_H, s_2 = s_H$
$R = R_L$	0.64	0.69	1
$R = R_H$	0.36	0.36	0

Table 2: **Acceptance probability given the outside option and the costs s_1 and s_2 when the attribute order is chosen optimally by the sales agent.**

This induces the consumer to learn both the first and the second attribute. Thus, the equilibrium probability that the high type accepts is given by 0.36. Table 2 presents the remaining probabilities for the various obfuscation choices of the firm. We can use this result to obtain the firm's profits. Denote with $\pi(s_1, s_2)$ the firm's profits as a function of the chosen costs s_1 and s_2. Then, we obtain with Table 2:

$$\pi(s_L, s_L) = 0.64 - 0.28\gamma$$
$$\pi(s_L, s_H) = 0.69 - 0.33\gamma$$
$$\pi(s_H, s_H) = 1 - \gamma.$$

One can see that the firm will never prefer to pick $s_1 = s_2 = s_L$ as this is dominated for all γ by the option of choosing $s_1 = s_L$ and $s_2 = s_H$. Moreover, when the fraction of consumers with a low outside option is rather large, that is γ is smaller than 0.46, the firm strictly prefers to set both costs to be high as this induces consumers with a low outside option to accept without learning. While high types with $R = R_H$ will reject the offer, their fraction is then sufficiently small so that it becomes profitable to simply ignore them.

4.2 Regulation of Information Provision

There are plenty of examples where the provision of information is regulated. Notably, this form of regulation not only considers which pieces of information should be provided but also in which format. For instance, the consumer credit directive (2008/48/EC) prescribes that lenders have to provide essential information to consumers in a standardized form. The following application formalizes a rationale for these considerations by showing that the order of information has implications for (consumer) welfare and should thus be considered when designing legislation.

Consider a market with a monopolistic seller who sells a product with three attributes v_1, v_2, and v_3 that are iid and follow the standard normal distribution. Learning costs are

$s_1 = 0.10$, $s_2 = 0.13$, and $s_3 = 0.18$. There are two types of consumers who differ in their outside options. In particular, $R_L = -0.5$ and $R_H = 0.5$ with $\Pr(R = R_H) = 0.5$. The rounded gross marginal benefits of learning for both types are displayed in Table 3. Suppose

Attribute i	1	2	3
MB_i	0.20	0.15	0.12

Table 3: **Rounded marginal benefits of learning with three iid standard normal attributes.**

that the seller learns the consumer's outside option when matched. When regulation simply requires the seller to make all information about the product accessible and consumers have learning costs, then the seller will optimally present the order $\sigma^* = (1, 3, 2)$ to the low type consumer with $R = R_L$, and the order $\sigma^* = (3, 2, 1)$ to the high type with $R = R_H$. In this case, the low type learns about one attribute and obtains an expected net surplus of 0.10. The high type learns about all 3 attributes and obtains in expectation 0.06. This is, however, not an optimal outcome from the consumers' perspective. If the regulator additionally stipulates that information must be presented in the order $\sigma^* = (1, 2, 3)$, then both types will learn about two attributes and both types obtain a net payoff of 0.12. Thus, as predicted by Proposition 1, in equilibrium there is an under-provision of information for the low type and an over-provision of information for the high type - even though consumers rationally decide how much to learn - which justifies the additional regulation addressing the format of information.

4.3 Search and Information Order

Lastly, I embed the (bilateral) baseline model described in Section 2 into a search market. This allows to endogenize the outside option R as the continuation value of search and to study how different types of information costs (search and learning costs) differ in the way they affect equilibrium outcomes. In particular, I consider a market with an infinite number of symmetric firms (senders) who offer products with N attributes and a single consumer (receiver) who searches sequentially at search costs $h \geq 0$ for matches and applies simultaneous learning when she learns about the match value of a particular firm's offer. The timing is as follows. First, firms determine the order in which the consumer can learn about their product's attributes. Then, the consumer can expend search costs h and be matched with a random firm. She then observes the order in which she can learn about the firm's product and decides how much to learn. If she decides to accept the firm's offer, the game ends. If she chooses to reject the offer, she re-enters the market and is matched - for costs h - with a new firm. There is no recall of past firms. I characterize the symmetric equilibrium of this search model in Appendix B.

I show that there is a fundamental difference in how search and learning costs affect the learning outcome, i.e., the number of learned attributes. The equilibrium outside option is decreasing in the consumer's search costs for a new firm. This implies, together with the inverse U-shape of the equilibrium learning decision in the outside option, that the learning decision is non-monotonic in the search costs. In contrast, the learning decision always decreases in the learning costs according to Corollary 2.

5 Discussion

Commitment to an Information Order Key assumptions of this paper are that, first, the sender commits ex-ante to an order in which information can be acquired by the receiver, and, second, the receiver actually adheres to the order chosen by the sender. The first assumption is rather innocuous, and numerous examples from the real world come to mind where this assumption holds. During a presentation, for instance, it is generally the presenter who controls the order in which different pieces of information are presented. In a marketing context, a consumer is likely to observe information in an advertisement first and only then she might see information provided on the firm's website or in its shop. In this setting, firms can clearly decide which information to convey over which channels and thus which order of information to choose.

One has to be more cautious with the second assumption. In some settings, e.g., during a presentation, a receiver has to listen to the information in the order chosen by the sender and she has no or only very limited opportunities to interact with the sender. In other cases, e.g. during a sales talk or while reading a product sheet, the receiver can try to proactively obtain pieces of information that are most relevant for her decision. However, in the first case, the sales agent is still in full control and might postpone answers to questions by the receiver to a later point in his sales talk. When information is presented in a written form and there is a simple way to navigate it, e.g., with an easily understandable table of contents, it clearly may be the case that this assumption is violated. Even then, it may still be the case that only experienced and sophisticated consumers are able to pick the most important pieces of information, while less sophisticated individuals will stick to the order provided by the sender.

Role of Receiver Privacy Another key assumption of this paper is that the sender is informed about the receiver's outside option R. In today's market place, notably in online markets, this assumption can be justified by referring to sellers' ability to track consumer behavior with tracking technologies like cookies. From a regulatory perspective thus the question arises whether a ban of such technologies, so that the sender no longer has information on the receiver's outside option, will ameliorate consumer welfare. The following example illustrates that such a regulation may backfire, and thus the regulator must pay close attention to the details of the marketplace.

Suppose the sender's offering has two attributes $v_i \sim \mathcal{N}(0, 1)$ with learning costs of $s_1 = 0.12$ and $s_2 = 0.18$. Learning is simultaneous and the outside option can be either low, $R_L = -0.5$, or high, $R_H = 0.5$, with $\Pr(R = R_H) = \gamma$. Then, it holds for both types of outside options that $MB_1 = 0.20$ and $MB_2 = 0.15$. If the sender observes the receiver's outside option, the expected receiver surplus is given by $0.06(1-\gamma) + 0.05\gamma$ and the sender plays $\sigma^* = (1, 2)$ when $R = R_L$ and $\sigma^* = (2, 1)$ when $R = R_H$. When the outside is unknown to the sender, he plays $\sigma^* = (1, 2)$ is $\gamma \leq 0.5$, and he plays $\sigma^* = (2, 1)$ when $\gamma > 0.5$. Thus, the receiver surplus is 0.05 when $\gamma > 0.05$ and it equals 0.06 when $\gamma \leq 0.05$. Therefore, a regulation that protects the receiver's privacy and thus effectively hides the receiver's outside option from the sender is beneficial only if the fraction of high types is sufficiently low. Note, however, that this conclusion depends on the details of the market. One can construct examples where such a privacy regulation is only beneficial if there is a sufficiently large fraction of receivers with a high outside option.[14]

6 Extension: Differences in Dispersions

One can readily argue that attributes in real world settings not necessarily differ in the costs that are necessary to understand them but rather in how polarizing they are. For

[14] This would be the case when $s_1 = 0.16$ and $s_2 = 0.21$ as in this case the sender would present a receiver with $R = R_L$ the order $\sigma^* = (2, 1)$ and one with $R = R_H$ the order $\sigma^* = (1, 2)$.

example, when considering a mobile phone, most people will have the same opinion on a specific resolution of the screen. In contrast, when the phone has an unusual color, e.g. pink, consumers will either strongly like or dislike this attribute.

Thus, within the simultaneous learning framework, I examine next the case when attributes v_i all have the same information acquisition costs, so that $s_i = s$, but differ in their dispersions. In particular, I impose that attributes v_i are members of a family of random variables $\{v^\theta\}$ with cdfs $F_\theta(v)$ and $E[v^\theta] = 0$ with the property that when $\theta_H > \theta_L > 0$ it holds that v^{θ_H} is a rotation of v^{θ_L} in the sense of Definition 1. Additionally, the respective cdfs $F_\theta(v)$ are continuous in θ and v and the rotations of the attributes $v_1, ..., v_N$ are given with $\theta_1 > \theta_2 > .. > \theta_N > 0$. One can interpret the rotation of an attribute as the degree for how polarizing it is.[15] In this section define $V(T, \sigma) = \sum_{i=1}^{T} v_{\sigma(i)}$ with $V(0, \sigma) = 0$. Denote with $H_{T,\sigma}(v)$ the respective cdf and with $MB_{i,\sigma} = E\left[\max\{R, V(i, \sigma)\}\right] - E\left[\max\{R, V(i-1, \sigma)\}\right]$ the marginal benefit of learning the i-th attribute when the order is σ. Note that in contrast to the baseline model, now the marginal benefits of learning are no longer independent of the order of attributes.

Analysis It is intuitive that the rotation of a sum of random variables increases if the rotation of any individual variable increases or when additional attributes are added to the sum. However, without further assumptions it is generally not possible to compare the rotation of two sums of members of such family of random variables. Therefore, I stipulate that the rotation $\Theta(T, \sigma)$ of the posterior expectation $V(T, \sigma)$ is given by the sum of the rotations of the individual attributes:

$$\Theta(T, \sigma) = \sum_{i=1}^{T} \theta_{\sigma(i)}.$$

This property holds for example when $\{v^\theta\}$ is a family of normal random variables with mean zero and variance $\mathrm{Var}(v^\theta) = \theta$. Recall that the sender's objective is $\Pi = 1 - H_{T,\sigma}(R)$. The sender's goal is thus to minimize (maximize) the rotation of the receiver's posterior expectation when the outside option is small (large) rather than simply minimize (maximize) the number of learned attributes. Denote with Θ^* the equilibrium rotation

[15] The rotation property has been frequently applied as a tool to model whether a product design is considered to appeal to a niche or a mass market, see, for example, Johnson and Myatt (2006) or Bar-Isaac et al. (2012).

that results from the sender picking an order that maximizes his payoff given the receiver's optimal behavior.

Surprisingly, when attributes differ in dispersions the sender optimal outcome can no longer be always achieved by an order that presents the most (or least) dispersed attributes first as the subsequent example shows. I present next an example where the order $\sigma = (2, 3, 1)$ is uniquely optimal. In what follows, v_i are normally distributed with mean zero and variances $\theta_1 = 4.77$, $\theta_2 = 3.28$, and $\theta_3 = 1.94$. Learning a single attribute costs $s = 0.34$ and suppose the outside option is $R = -0.5$. The net values of learning T attributes for all possible orders σ as well as the respective optimum $T^*(\sigma)$ are displayed in Table 4.

	$T = 1$	$T = 2$	$T = 3$	$T^*(\sigma)$
$\sigma = (1, 2, 3)$	0.3018	0.2154	0.001	1
$\sigma = (1, 3, 2)$	0.3018	0.1189	0.001	1
$\sigma = (2, 1, 3)$	0.1582	0.2154	0.001	2
$\sigma = (2, 3, 1)$	0.1582	0.0000	0.001	1
$\sigma = (3, 1, 2)$	0.0000	0.1189	0.001	2
$\sigma = (3, 2, 1)$	0.0000	0.0000	0.001	3

Table 4: **The receiver's net surplus from learning T attributes net of costs and the receiver optimal learning decision $T^*(\sigma)$.**

The sender's objective is to minimize the rotation of the receiver's posterior expectation $V(T^*(\sigma), \sigma)$. When attributes differ in their dispersion, the number of attributes learned is only an imperfect proxy for the dispersion of $V(T^*(\sigma), \sigma)$ as now not only the number of learned attributes matters but also their respective levels of dispersion θ_i. In Table 4 we see that orders $\sigma = (1, 2, 3)$, $\sigma = (1, 3, 2)$, and $\sigma = (2, 3, 1)$ all result in the same number of attributes learned: $T^* = 1$. However, the sender will strictly prefer the order $\sigma = (2, 3, 1)$ as this order minimizes the resulting rotation of the posterior expectation. The next proposition summarizes the results from this model variant. The comparative results follow the same intuition as in the baseline model in Section 2.

Proposition 3 *When the receiver learns simultaneously about an offer in an order chosen by the sender, and attributes are heterogeneous in dispersions, the equilibrium is characterized as follows. The sender chooses an order that minimizes (maximizes) the rotation of posterior expectations Θ^* when the receiver's outside is small with $R < 0$ (large with*

$R > 0$). Moreover, the equilibrium dispersion Θ^* is increasing in R when $R < 0$, and it is decreasing in R when $R > 0$. The equilibrium level of dispersion decreases in the learning costs s.

As it is not possible to determine the shape of optimal orders, it is also not possible to characterize in general how the distribution of dispersions across attributes affects equilibrium outcomes. Still, it continues to hold that the distribution has an impact. Consider the following example. Suppose the offering has two normally distributed attributes v_1 and v_2 with $E[v_i] = 0$ and their variances are given by θ_i. Learning costs are $s = 0.08$ and the receiver's outside option is $R = -0.5$. Table 5 shows the gross marginal benefits of learning the first and the second attribute for the various orders, first for the case when $\theta_1 = 0.55$ and $\theta_2 = 0.45$ and then for the case when $\theta_1 = 0.7$ and $\theta_2 = 0.3$. The receiver learns both attributes in the first case irrespective of the order. Thus, the equilibrium dispersion is given by $\theta_1 + \theta_2 = 1$. However, when the costs are sufficiently concentrated as in the second case, the sender can induce the receiver to learn only one attribute when the order is $\sigma = (1, 2)$ even though the dispersion of the sum of attributes remains constant. Note that a further increase in θ_1 at the expense of θ_2 (so that still $\theta_1 + \theta_2 = 1$) will decrease the sender's payoffs.

	$MB_{1,\sigma}$	$MB_{2,\sigma}$
$\sigma = (1, 2)$	0.1107	0.0871
$\sigma = (2, 1)$	0.0887	0.1091

	$MB_{1,\sigma}$	$MB_{2,\sigma}$
$\sigma = (1, 2)$	0.1417	0.0561
$\sigma = (2, 1)$	0.0537	0.1441

Table 5: **Marginal benefits of learning when $\theta_1 = 0.55$ and $\theta_2 = 0.45$ (left table) and when $\theta_1 = 0.7$ and $\theta_2 = 0.3$ (right table).**

Further Extensions I consider an additional extension in Appendix B which revisits the case of dispersion heterogeneity when infinitely many attributes determine the value of the offering so that the impact of the individual attribute becomes arbitrarily small. Interestingly, in this limit case, I am able to fully characterize the sender optimal order and the impact of the distribution of dispersions analogously to Corollary 4.

7 Conclusion

When learning is costly, the order in which various pieces of information can be acquired has a crucial impact on which pieces of information are learned and how much information is acquired. A sender who has the ability to control the sequence of information can use this mechanism in order to persuade even a receiver who is otherwise fully rational. In this paper, I study a model where a sender has a multi-attribute offering and can determine the order in which information is acquired by a receiver. This paper characterizes the optimal orders of information provision in various settings and highlights the role of the distribution of processing costs across attributes as a key determinant for outcomes.

The results in this paper have important implications for businesses and consumer protection. For business and marketing purposes, this paper provides a guideline for optimal presentation of information and reveals that (ex-ante) design and obfuscation choices must be made under the consideration of (ex-post) communication with optimal information orders. From the perspective of regulation, this paper shows that simply regulating the supply of information provided by firms can be considered an insufficient measure for consumer protection. Even when the content of information provision is regulated, firms still have plenty of room to persuade consumers. In particular, those consumers who are less sophisticated, and thus are less likely to choose information optimally, may fall prey to persuasion through ordered information.

Regarding future research, I see multiple exciting directions. While the present paper deals with a single receiver, it is of interest how the persuasion of an audience should be treated, for example how should the arguments in a press conference of a central bank be organized. The results of this paper extend to audiences where the outside options of its members are rather concentrated, however, outcomes are no longer clear in cases when the outside options are rather dispersed. Second, one should consider settings where the sender has private information about certain attributes. Third, it is important to cover situations where the receiver plays a more active role. During a sales pitch, for instance, a receiver could request a particular piece of information. Lastly, it is of interest for applications, such as the design of an online shop, to study the optimal ordering of information in the context of a sender who has multiple offers with multiple attributes.

References

Arbatskaya, M. (2007). Ordered search. *The RAND Journal of Economics 38*(1), 119–126.

Armstrong, M. (2017). Ordered consumer search. *Journal of the European Economic Association 15*(5), 989–1024.

Armstrong, M., J. Vickers, and J. Zhou (2009). Prominence and consumer search. *The RAND Journal of Economics 40*(2), 209–233.

Baird, J. and R. C. Zelin (2000). The effects of information ordering on investor perceptions: An experiment utilizing presidents' letters. *Journal of Financial and Strategic Decisions 13*(3), 71–80.

Bar-Isaac, H., G. Caruana, and V. Cuñat (2012). Search, design, and market structure. *The American Economic Review 102*(2), 1140–1160.

Bergus, G. R., G. B. Chapman, B. T. Levy, J. W. Ely, and R. A. Oppliger (1998). Clinical diagnosis and the order of information. *Medical Decision Making 18*(4), 412–417. PMID: 10372584.

Branco, F., M. Sun, and J. M. Villas-Boas (2012). Optimal search for product information. *Management Science 58*(11), 2037–2056.

Chade, H. and E. Schlee (2002). Another look at the Radner-Stiglitz nonconcavity in the value of information. *Journal of Economic Theory 107*(2), 421 – 452.

Dominique-Ferreira, S. (2017). How important is the strategic order of product attribute presentation in the non-life insurance market? *Journal of Retailing and Consumer Services 34*, 138–144.

Ellison, G. and A. Wolitzky (2012). A search cost model of obfuscation. *The RAND Journal of Economics 43*(3), 417–441.

Gamp, T. (2017). Guided search. *Working paper*.

Ganuza, J.-J. and J. S. Penalva (2010). Signal orderings based on dispersion and the supply of private information in auctions. *Econometrica 78*(3), 1007–1030.

Hellmann, A., C. Yeow, and L. De Mello (2017). The influence of textual presentation order and graphical presentation on the judgements of non-professional investors. *Accounting and Business Research 47*(4), 455–470.

Hoffmann, F. and R. Inderst (2011). Pre-sale information. *Journal of Economic Theory 146*(6), 2333–2355.

Hogarth, R. M. and H. J. Einhorn (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology 24*(1), 1–55.

Johnson, J. P. and D. P. Myatt (2006). On the simple economics of advertising, marketing, and product design. *The American Economic Review 96*(3), 756–784.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Klabjan, D., W. Olszewski, and A. Wolinsky (2014). Attributes. *Games and Economic Behavior 88*, 190–206.

Levav, J., M. Heitmann, A. Herrmann, and S. S. Iyengar (2010). Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy 118*(2), 274–299.

Purkayastha, S. (1998). Simple proofs of two results on convolutions of unimodal distributions. *Statistics & Probability Letters 39*(2), 97–100.

Radner, R. and J. Stiglitz (1984). A nonconcavity in the value of information. *Bayesian models in economic theory 5*, 33–52.

Vives, X. (2000). *Oligopoly pricing: Old ideas and new tools.* MIT Press.

A Appendix: Proofs

Proof of Lemma 1. Consider two random variables v_1 and v_2 and suppose that they are independently distributed according to $g_1(\cdot)$ and $g_2(\cdot)$ with cdf $G_1(\cdot)$ and $G_2(\cdot)$, respectively. Suppose further that both densities are symmetric and unimodal around mean 0. Then the density of their sum, $v_1 + v_2$, is given by the convolution $h(\cdot)$ with

$$h(v) = \int_{-\infty}^{\infty} g_1(v_1)g_2(v - v_1)dv_1 = \int_{-\infty}^{\infty} g_2(v_2)g_1(v - v_2)dv_2$$

which is unimodal and symmetric around 0 as well, see Theorem 2.1 in Purkayastha (1998), which implies that $V(T)$ is unimodal and symmetric for all T. Denote the cdf of the sum $v_1 + v_2$ with $H(\cdot)$. I show that for $R < 0$ it holds that $H(R) > G_i(R)$ and for $R > 0$ it holds that $H(R) < G_i(R)$ for $i = 1, 2$ which implies that $v_1 + v_2$ is obtained by a rotation from both v_1 and v_2. The cdf of the sum is given by

$$H(R) = \int_{-\infty}^{\infty} G_1(R - v_2)dG_2(v_2) = \int_{-\infty}^{\infty} G_2(R - v_1)dG_1(v_1).$$

I am interested in the sign of $H(R) - G_2(R)$ so

$$\begin{aligned} H(R) - G_2(R) &= \int_{-\infty}^{\infty} G_1(R - v_2) - \mathbf{1}_{v_2 \le R} dG_2(v_2) \\ &= \int_{0}^{\infty} [1 - G_1(R + x)] (g_2(R + x) - g_2(R - x)) dx. \end{aligned}$$

Due to the unimodality and symmetry around 0 of $g_2(\cdot)$, clearly $H(R) - G_2(R) > 0$ when $R < 0$ and $H(R) - G_2(R) < 0$ when $R > 0$. The claim for $H(R) - G_1(R)$ is analogous. **Q.E.D.**

Proof of Lemma 2. Suppose when the order is σ, the receiver's payoff is maximized by $T^*(\sigma)$. When $T^* > 0$, define the order σ' as a copy of σ where the attributes v_N and $v_{T^*(\sigma)}$ are swapped. Then with $s_N > \max_i\{MB_i\} \geq MB_{T^*(\sigma)}$ it is no longer optimal for the receiver to learn $T^*(\sigma)$ attributes. When $T^*(\sigma) = 0$, construct the order σ' from the order σ by swapping attributes v_1 and $v_{\sigma(1)}$, and the receiver will prefer learning at least one attribute to not learning at all. **Q.E.D.**

Proof of Lemma 3. Consider an order σ with $T^*(\sigma) = k$. I show that with the order σ_k

it is optimal for the receiver to learn k attributes. Suppose first that $l < k$. Then,

$$\sum_{i=1}^{k} \left(MB_i - s_{\sigma_k(i)}\right) - \sum_{i=1}^{l} \left(MB_i - s_{\sigma_k(i)}\right) = \sum_{i=l+1}^{k} \left(MB_i - s_{k+1-i}\right) \geq \sum_{i=l+1}^{k} \left(MB_i - s_{\sigma(i)}\right) \geq 0$$

where the first inequality follows from $\sum_{i=l+1}^{k} s_{k+1-i} = \sum_{i=1}^{k-l} s_i \leq \sum_{i=l+1}^{k} s_{\sigma(i)}$ and the second from the optimality of $k = T^*(\sigma)$ when the order is σ. Similarly, when $l > k$ it holds

$$\sum_{i=1}^{k} \left(MB_i - s_{\sigma_k(i)}\right) - \sum_{i=1}^{l} \left(MB_i - s_{\sigma_k(i)}\right) = -\sum_{i=k+1}^{l} \left(MB_i - s_{N+k+1-i}\right) \geq -\sum_{i=k+1}^{l} \left(MB_i - s_{\sigma(i)}\right) \geq 0$$

where the first inequality follows from $\sum_{i=k+1}^{l} s_{N+k+1-i} = \sum_{i=N+k+1-l}^{N} s_i \geq \sum_{i=k+1}^{l} s_{\sigma(i)}$ and the second follows again from the optimality of $k = T^*(\sigma)$. Clearly, when $k = 0$ or $k = N$, then only one of the discussed cases applies. **Q.E.D.**

Proof of Lemma 4. I show that, given that the receiver chooses optimally how many attributes to learn, the order $\sigma^{**}(i) = i$ results in higher payoffs for the receiver than all orders $\sigma \in \mathcal{S}$. Notably, it must hold

$$\begin{aligned}
&E\left[\max\{R, V(T^*(\sigma^{**}))\}\right] - \sum_{i=1}^{T^*(\sigma^{**})} s_{\sigma^{**}(i)} \\
\geq &E\left[\max\{R, V(T^*(\sigma))\}\right] - \sum_{i=1}^{T^*(\sigma)} s_{\sigma^{**}(i)} \\
\geq &E\left[\max\{R, V(T^*(\sigma))\}\right] - \sum_{i=1}^{T^*(\sigma)} s_{\sigma(i)}
\end{aligned}$$

where the first inequality follows from the optimality of $T^*(\sigma^{**})$ when the order is σ^{**}, and the second one follows from $\sum_{i=1}^{T^*(\sigma)} s_{\sigma(i)} \geq \sum_{i=1}^{T^*(\sigma)} s_i$. **Q.E.D.**

Proof of Proposition 1. With Lemma 1 and optimality, the sender chooses an order that implements a learning decision $k^* = \min\{\mathcal{K}_R\}$ when $R < 0$ and $k^* = \max\{\mathcal{K}_R\}$ when $R > 0$. With Lemma 3, it follows that σ_{k^*} implements this learning decision. The results regarding the under- and over-provision of information follow from the discussion in the main text preceding Proposition 1. **Q.E.D.**

Proof of Lemma 5. It holds that

$$\frac{\partial MB_i}{\partial R} = H_i(R) - H_{i-1}(R)$$

which is with Lemma 1 and Definition 1 positive when $R < 0$ and negative when $R > 0$ for all $T = 1, ..., N$. **Q.E.D.**

Proof of Corollary 1. For any order $\sigma \in \mathcal{S}$, it holds with Lemma 5 that gross marginal benefits of learning MB_i are increasing in $R < 0$ and decreasing in $R > 0$. This implies with arguments from monotone comparative statics, cf. Theorem 2.3 in Vives (2000), that the set of solutions to the receiver problem is strictly increasing in $R < 0$ and strictly decreasing in $R > 0$ for all orders $\sigma \in \mathcal{S}$. It follows that $\min\{\mathcal{K}_R\}$ increases in R when $R < 0$, and $\max\{\mathcal{K}_R\}$ decreases in R when $R > 0$. **Q.E.D.**

Proof of Corollary 2. Standard arguments from monotone comparative statics, cf. Theorem 2.3 in Vives (2000), imply that when the cost s_i increases by $\Delta s > 0$, the set of solutions of the receiver problem decreases for all $\sigma \in \mathcal{S}$, and the first claim follows.

For the second claim, consider a cost shift such that s_i decreases by $\Delta s > 0$ and s_j increases by Δs with $i < j$. I show first that when a learning decision k was implementable before the cost shift, it remains implementable after the cost shift. There are three cases to consider. Suppose first that $i < j \leq k$. Then, the receiver's expected payoff from learning l attributes following the order σ_k is given by

$$E\left[\max\{R, V(l)\}\right] - \sum_{i=1}^{l} s_{\sigma_{k^*}(i)},$$

and it decreases when $k+1-j \leq l < k+1-i$ while the expected payoff from any other learning decision remains constant. When $k+1 \leq i < j$, the expected receiver payoff from learning l attributes following σ_k decreases when $N+k+1-j \leq l < N+k+1-i$ and remains constant otherwise. Lastly, when $i \leq k \leq j$, the expected payoff from learning $k+1-i \leq l < N+k+1-j$ following σ_k increases by Δs. In particular, also the payoff from learning k attributes increases by Δs. The payoffs from the remaining learning decisions remain constant. Taken together, it follows that the learning decision k remains implementable.

Second, I show that there are cost shifts that allow the sender to implement learning decisions that were not implementable before the shift. Suppose, for instance, $R < 0$ and $k^* = \min\{\mathcal{K}_R\} \geq 2$ and the shift is given with $i < j \leq k^*$ and $s_j + \Delta s > MB_{k^*}$. After such a cost shift, clearly, the sender can implement a learning decision that is smaller than k^*. **Q.E.D.**

Proof of Lemma 6. Consider first the case when search costs $s_{\sigma(2)}$ are so high that learning the second attribute is never optimal. Then, the receiver accepts the outside option only if $R > v_{\sigma(1)}$ and, therefore,

$$\underline{b} = \overline{b} = R.$$

Next, consider the case when search costs $s_{\sigma(2)}$ are sufficiently small so that learning is strictly preferred for some realizations v of $v_{\sigma(1)}$. With the Leibniz's rule, I obtain

$$\frac{\partial}{\partial v} E\left[\max\{R, v + v_{\sigma(2)}\}\right] = 1 - F(R - v)$$

so that the gross value of learning is increasing in v. This implies together with $\lim_{v \to -\infty} E[\max\{R, v + v_{\sigma(2)}\}] = R$ that there exists a unique point $\underline{b}$ such that

$$R = E\left[\max\{R, \underline{b} + v_{\sigma(2)}\}\right] - s_{\sigma(2)}, \tag{8}$$

where the receiver is indifferent between the outside option and learning. Moreover, the slope of the value of learning in v implies that there is a unique value $\overline{b}$ such that

$$\overline{b} = E\left[\max\{R, \overline{b} + v_{\sigma(2)}\}\right] - s_{\sigma(2)} \tag{9}$$

which determines the receiver who is indifferent between learning and accepting the offering.

It remains to show that $\underline{b} < \overline{b}$. For this, I show that $\underline{b} < R$ and $R < \overline{b}$. Suppose, for a contradiction, that $\underline{b} > R$. Then learning could not be optimal for any realization v of $v_{\sigma(1)}$ as it holds for $v > R$ that accepting is strictly better than the outside option, and the value of accepting the offering increases at a higher rate in $v = v_{\sigma(1)}$ than the value of learning. Next, suppose that $\overline{b} < R$. Then, I obtain again a contradiction since the value of accepting increases at a higher rate in $v = v_{\sigma(1)}$ than the value of learning, and both are below R at $\overline{b}$ so that learning is never optimal. **Q.E.D.**

Proof of Lemma 7. Consider first the case when $R \leq 0$. Clearly, $L_1 + s_{\sigma(1)} > 0$ so that when $s_{\sigma(1)}$ is sufficiently small, learning is strictly preferred to accepting the offering. The case when $R > 0$ follows by the same arguments. **Q.E.D.**

Proof of Lemma 8. Differentiation of the sender's payoffs with respect to s results in

$$\frac{d\Pi}{ds} = -f(\bar{b})\frac{d\bar{b}}{ds} + [1 - F(R - \bar{b})]f(\bar{b})\frac{d\bar{b}}{ds} - [1 - F(R - \underline{b})]f(\underline{b})\frac{d\underline{b}}{ds}.$$

When learning costs s are not too high, I obtain with the implicit function theorem and the equations (8) and (9) that

$$\frac{d\underline{b}}{ds} = \frac{1}{1 - F(R - \underline{b})}$$
$$\frac{d\bar{b}}{ds} = -\frac{1}{F(R - \bar{b})},$$

and substitution yields

$$\frac{d\Pi}{ds} = f(\bar{b}) - f(\underline{b}). \tag{10}$$

In order to sign the expression in (10), I show first that $R - \underline{b} = \bar{b} - R$ for all s. Denote with $\Delta_0(v) \equiv E[\max\{R, v + v_{\sigma(2)}\}] - s - R$ and $\Delta_1(v) \equiv E[\max\{R, v + v_{\sigma(2)}\}] - s - v$. By definition, it holds that $\Delta_0(\underline{b}) = 0$ and $\Delta_1(\bar{b}) = 0$. Then, it holds that

$$\left.\frac{\partial \Delta_0}{\partial v}\right|_{R-w} = -\left.\frac{\partial \Delta_1}{\partial v}\right|_{R+w}$$

for all $w \geq 0$ as we have that

$$\left.\frac{\partial \Delta_0}{\partial v}\right|_{R-w} = 1 - F(w)$$
$$\left.\frac{\partial \Delta_1}{\partial v}\right|_{R+w} = -F(-w),$$

and that $1 - F(w) = F(-w)$ due the symmetry of $f(\cdot)$ around mean 0. Since $\Delta_0(R) = \Delta_1(R)$, the difference between $\underline{b}$ and R is the same as the difference between $\bar{b}$ and R. From this follows, together with the unimodality and symmetry of $f(\cdot)$, that $d\Pi/ds > 0$ when $R < 0$ and $d\Pi/ds < 0$ when $R > 0$. **Q.E.D.**

Proof of Proposition 2. Consider first the case when $R < 0$, and express the value of learning in $t = 1$ for a given order σ as $L_1(\sigma)$. Then the receiver will clearly not learn if $L_1((2,1)) \leq 0$ since accepting results in a higher expected payoff. This implicitly defines $a^-(s_1)$ as the learning costs for the attribute v_2 that result in $L_1((2,1))|_{s_2=a^-(s_1)} = 0$. Moreover, we have that

$$\begin{aligned}\frac{dL_1((2,1))}{ds_1} &= Rf(\underline{b})\frac{d\underline{b}}{ds_1} - \left(F(\overline{b}) - F(\underline{b})\right) + \left(E\left[\max\{R, \overline{b} + v_1\}\right] - s_1\right) f(\overline{b})\frac{d\overline{b}}{ds_1} \\ &- \left(E\left[\max\{R, \underline{b} + v_1\}\right] - s_1\right) f(\underline{b})\frac{d\underline{b}}{ds_1} - \overline{b}f(\overline{b})\frac{d\overline{b}}{ds_1} \\ &= -\left(F(\overline{b}) - F(\underline{b})\right) < 0\end{aligned}$$

where the last equality follows from the fact that by definition $R = E\left[\max\{R, \underline{b} + v_1\}\right] - s_1$ and $\overline{b} = E\left[\max\{R, \overline{b} + v_1\}\right] - s_1$. This implies that $a^-(s_1)$ is decreasing in s_1. Thus, when $s_2 \geq a^-(s_1)$ the sender optimally chooses $\sigma = (2,1)$ as then the receiver always accepts the offering. Next, consider the case when $s_2 < a^-(s_1)$. Then, the receiver learns in $t = 1$ given the order $\sigma = (2,1)$. I show next that if the receiver starts learning with the order $\sigma = (2,1)$ then she will strictly prefer to do so given the order $\widehat{\sigma} = (1,2)$. Denote the respective optimal cutoffs by $\underline{b}(\sigma)$, $\overline{b}(\sigma)$, $\underline{b}(\widehat{\sigma})$, and $\overline{b}(\widehat{\sigma})$. Then, I obtain

$$\begin{aligned}&L_1(\widehat{\sigma}) - L_1(\sigma) = \\ &F(\underline{b}(\widehat{\sigma}))R + \int_{\underline{b}(\widehat{\sigma})}^{\overline{b}(\widehat{\sigma})} E\left[\max\{R, v + v_2\}\right] - s_2 dF(v) + \int_{\overline{b}(\widehat{\sigma})}^{\overline{v}_{\widehat{\sigma}(1)}} v dF(v) - s_1 \\ &-F(\underline{b}(\sigma))R - \int_{\underline{b}(\sigma)}^{\overline{b}(\sigma)} E\left[\max\{R, v + v_2\}\right] - s_1 dF(v) - \int_{\overline{b}(\sigma)}^{\overline{v}_{\sigma(1)}} v dF(v) + s_2 \geq \\ &(s_2 - s_1) - \int_{\underline{b}(\sigma)}^{\overline{b}(\sigma)} (s_2 - s_1) dF(v) \geq 0\end{aligned}$$

where I use for the first inequality the fact that with the order $\widehat{\sigma}$ the cutoffs $\underline{b}(\sigma)$ and $\overline{b}(\sigma)$ are not optimal and the fact that the distributions of v_1 and v_2 are identical. This, together with Lemma 8, implies that the sender prefers $\sigma = (1,2)$ over $\sigma = (2,1)$, and the receiver learns the realization of at least one attribute.

Next, consider the case when $R > 0$. From Lemma 8 it follows that when $R > 0$, the sender prefers, conditional on the receiver learning the first attribute, that the learning costs of the second attribute are as low as possible. Thus, the sender prefers the order $\sigma = (2,1)$ whenever s_2 is sufficiently low so that $L_1((2,1)) \geq R$. By the same arguments as above, $L_1((2,1))|_{s_2=a^+(s_1)} = R$ defines the decreasing function $a^+(s_1)$ so that when $s_2 \leq a^+(s_1)$ the sender chooses $\sigma = (2,1)$, and the receiver learns at least one attribute. If $s_2 > a^+(s_1)$, the sender weakly prefers $\sigma = (1,2)$. Notably, if costs s_1 and s_2 are sufficiently low so that $L_1((2,1))|_{s1,s2} \geq R$, the receiver learns at least one attribute. **Q.E.D.**

Proof of Corollary 3. I first show that for any order σ there are cutoffs $\underline{R} < 0 < \overline{R}$ so that the probability to learn one attribute equals 1 when $\underline{R} < R \leq \overline{R}$ and 0 otherwise. For both orders $\sigma = (1,2)$ and $\sigma = (2,1)$ the probability of learning one attribute increases in $R < 0$, and it decreases in $R > 0$. When $R < 0$, the receiver will either learn or accept the offering. As we have

$$\frac{dL_1}{dR} = F(\underline{b}) + \int_{\underline{b}}^{\bar{b}} F(R-v)dF(v) > 0,$$

the net value of learning is increasing in R when $R < 0$. When $R > 0$, the receiver will either learn or take her outside option. As it holds that

$$\frac{d}{dR}\left[L_1 - R\right] = F(\underline{b}) + \int_{\underline{b}}^{\bar{b}} F(R-v)dF(v) - 1 < 0,$$

the net value of learning is decreasing, and so it is highest when $R = 0$. With the assumption that there are some R where learning is optimal, the claim follows.

Next, I show for both orders that the probability of learning two attributes increases in the outside R when $R < 0$, and it decreases in R when $R > 0$. Clearly, the receiver can only learn the second attribute if she starts learning. Conditional on this, the probability that she learns the second attribute is given by $F(\bar{b}) - F(\underline{b})$ and so

$$\frac{d}{dR}\left[F(\bar{b}) - F(\underline{b})\right] = f(\bar{b}) - f(\underline{b})$$

since it holds that $d\underline{b}/dR = d\bar{b}/dR = 1$ so that, together with the unimodality and symmetry of $f(\cdot)$ around 0, the above expression is strictly positive when $R < 0$ and strictly negative when $R > 0$, and the claim follows.

With the result on the probability of learning one attribute and Proposition 2, in equilibrium, there are cutoffs $R' < 0$ and $R'' > 0$ such that the sender optimally chooses $\sigma = (2,1)$ when $R < R'$ and when $0 < R < R''$. Otherwise, when $R' < R < 0$ and when $R > R''$, the sender chooses $\sigma = (1,2)$, cf. Lemma 8. This proves the claim for the probability of learning one attribute. Lastly, note that the probability to learn two attributes has clearly a positive jump discontinuity at R'. It has a negative jump discontinuity at R'' since at this point the sender changes the order from $\sigma = (2, 1)$ to $\sigma = (1, 2)$ which clearly results in a discrete change in $\underline{b}$ and $\bar{b}$. **Q.E.D.**

Proof of Corollary 4. For the first claim, note that when the cost of learning s_i increases, the incentives to learn in any order decrease and so do the probabilities to learn one or two attributes.

For the second claim, it holds that $d\Pi/ds_{\sigma(2)} \geq 0$ when $R < 0$. Further, the payoff Π is constant in $s_{\sigma(1)}$ almost everywhere with the exception of a single positive jump discontinuity in $s_{\sigma(1)}$ at which $\Pi^* = 1$, cf. Lemma 7. This proves the first half of the second part.

Next, turn to the case with $R > 0$. When $s < \bar{s}$ where $\bar{s}$ is defined by $0 = L_1|_{s_{\sigma(1)}=\bar{s}, s_{\sigma(2)}=0}$, the sender's payoff increases when s_2 increases as then he will optimally choose $\sigma = (2, 1)$. Thus, learning costs $s_{\sigma(2)} = s_1$ decrease which increases the sender's expected payoff Π, cf. Lemma 8. When $s > \bar{s}$, there are two additional cases to distinguish. First, when $\bar{s} < s \leq \bar{\bar{s}}$ where $\bar{\bar{s}}$ is defined as the solution to $0 = L_1|_{s_{\sigma(1)}=s_{\sigma(2)}=\bar{\bar{s}}/2}$, the sender initially chooses the order $\sigma = (2, 1)$ so that his payoff is increasing in s_2. Note, however, that when $s_{\sigma(1)} = s_2$ and $s_{\sigma(2)} = s - s_2$ it holds that

$$\frac{d}{ds_2} L_1((2, 1)) = -1 + \big(F(\bar{b}) - F(\underline{b})\big) < 0$$

so that as s_2 increases, the sender has to switch the order at some point to $\sigma = (1, 2)$. After that point his payoffs decrease in s_2, cf. Lemma 8.

When $s > \bar{\bar{s}}$, the sender always chooses optimally the order $\sigma = (1, 2)$. As s_2 increases, $s_{\sigma(1)} = s - s_2$ decreases and so there is a positive jump discontinuity in the sender's payoffs as eventually the receiver learns at least one attribute. After that the sender's payoff weakly decreases in s_2 as this increases the costs of learning the second attribute. **Q.E.D.**

Proof of Proposition 3. The first part of the Proposition follows by the arguments in the main text.

For the second part, note that for any order $\sigma \in \mathcal{S}$ it holds that

$$\frac{\partial MB_{i,\sigma}}{\partial R} = H_{i,\sigma}(R) - H_{i-1,\sigma}(R)$$

which is positive for $R < 0$ and negative for $R > 0$ so that the receiver payoff has increasing differences in R and T when $R < 0$ and decreasing differences when $R > 0$. This shows by standard arguments from monotone comparative statics that the set of maximizers T^* is increasing in R for $R < 0$ and decreasing for $R > 0$ and implies that the resulting rotation for any order σ is, therefore, increasing in R when $R < 0$ and decreasing in R when $R > 0$. Since for $R < 0$ all rotations that the sender can implement are increasing this also holds for the smallest implementable rotation, and when $R > 0$, the rotations for all orders σ are decreasing and therefore also for the largest implementable rotation.

The last part of the Proposition follows by the same arguments as the Proof of Corollary 2. **Q.E.D.**

B Appendix: Additional Material

Search and Information Order

In this Section, I provide a characterization of the search market as described in the last part of Section 4. In equilibrium, the outside option R^* given as the continuation value of search satisfies

$$h = E\left[\max\{R, V(T^*)\}\right] - \sum_{i=0}^{T^*} s_{\sigma^*(i)} - R^* \tag{11}$$

where σ^* denotes here the equilibrium order of attributes with $s_{\sigma(0)} \equiv 0$, that is, given R^* the sequence σ^* solves the individual firm's problem of maximizing the likelihood of accepting given that the consumer's best response is given by T^*. Intuitively, on the left side of equation (11) are the costs and on the right side are the expected marginal benefits of one additional search when the (expected) value of the current offer is R^*. Now, the main primitive of interest are the search costs $h \geq 0$. Note that due to the discreteness of the pure strategies of the sender and the receiver there will be values of h such that the equation (11) has more than one solution in R. I impose that whenever this is the case the consumer believes that the smallest R determines his outside option and this belief is known to all firms in the market. Moreover, firms use only strategies from the set $\{\sigma_k\}_{k=1,\ldots,N}$. In order to avoid additional case distinctions and as to obtain a complete picture of comparative statics, I impose in this section only that $E[\sum v_i] = \mu > 0$ and learning costs are such that the consumer learns at least one attribute when $R = 0$. Then, I obtain the following result.

Proposition 4 *Suppose the consumer searches sequentially for firms and learns simultaneously about a firm's offer, while firms choose the order in which the consumer can learn. For any search cost $h \geq 0$ a unique equilibrium consumer continuation value $R^* \geq 0$ exists, and it is decreasing in h.*

Proof. I show first that the right side of (11) given that both the firm and the consumer behave optimally for given R is strictly decreasing in R. Note first that in equilibrium T^* will take on only a finite number of values. This implies that the real line can be split into half-open intervals of the form $(R_0, R_1]$ on which T^* is constant. On such an interval, the right side of equation (11) is strictly decreasing as it holds that

$$\frac{\partial}{\partial R}\left[E\left[\max\{R, V(T)\}\right] - \sum_{i=1}^{T} s_{\sigma(i)} - R\right] = H_T(R) - 1$$

for any order σ and any $1 \leq T \leq N$. When $T = 0$, that derivative holds only almost everywhere since there is a single point at $R = \mu$ where the right side of equation (11) is not differentiable but still continuous in R. Suppose l and m are the equilibrium numbers of learned attributes over the intervals $(R_0, R_1]$ and $(R_1, R_2]$, respectively, with $R_0 < R_1 < R_2$. I show next that

$$\lim_{R \to R_1^-}\left(E\left[\max\{R, V(T^*)\}\right] - \sum_{i=1}^{T^*} s_{\sigma^*(i)}\right) \leq \lim_{R \to R_1^+}\left(E\left[\max\{R, V(T^*)\}\right] - \sum_{i=1}^{T^*} s_{\sigma^*(i)}\right) \tag{12}$$

where T^* is optimally chosen by the consumer given the order σ^*. Consider first the case $R_0 < R < R_1 < \mu$ so that $l < m$, from the inverse U-shape of the learning decision around the mean. As R increases, all marginal benefits MB_i increase so that given the order σ_l the consumer is indifferent between learning l and $l' > l$ attributes at R_1. If not, the firm would still be able to implement learning of l attributes for some $R > R_1$. It must hold that $m \leq l'$. Due to the indifference at R_1, it holds

$$\sum_{i=1}^{l}(MB_i - s_{\sigma_l(i)}) = \sum_{i=1}^{l'}(MB_i - s_{\sigma_l(i)})$$

With the order σ_m, it must hold that

$$\begin{aligned}
&\sum_{i=1}^{m}(MB_i - s_{\sigma_m(i)}) - \sum_{i=1}^{l}(MB_i - s_{\sigma_l(i)}) \\
\geq &\sum_{i=1}^{l'}(MB_i - s_{\sigma_k(i)}) - \sum_{i=1}^{l'}(MB_i - s_{\sigma_l(i)}) \\
= &\sum_{i=1}^{l'-l} S_{N+1-i} - \sum_{i=1}^{l'-m} S_{N+1-i} - \sum_{i=1}^{m-l} s_{m+1-i} > 0,
\end{aligned}$$

where the first inequality follows from the optimality of m when σ_m the order faced by the consumer so that the above claim in (12) must hold. Next, consider the case when $\mu < R_0 < R < R_1$ so that $l > m$ with the inverse U-shape of the learning decision. As R increases, the marginal benefits of learning MB_i decrease so that at R_1 the consumer must be indifferent between learning l and $l' < l$ attributes given the order σ_l and when the order is σ_m she must have a weak preference for learning $m \geq l'$ attributes. Thus, we obtain again

$$\begin{aligned}
&\sum_{i=1}^{m}(MB_i - s_{\sigma_m(i)}) - \sum_{i=1}^{l}(MB_i - s_{\sigma_l(i)}) \\
\geq &\sum_{i=1}^{l'}(MB_i - s_{\sigma_k(i)}) - \sum_{i=1}^{l'}(MB_i - s_{\sigma_l(i)}) \\
= &\sum_{i}^{l'}(s_{l+1-i} - s_{m+1-i}) > 0.
\end{aligned}$$

This establishes that the right side of (11) is almost everywhere continuous and strictly decreasing in R with a finite number of positive jump discontinuities. By assumption, it holds that at $R = 0$ we have $T^*|_{R=0} \geq 1$ so that

$$\overline{h} \equiv \left.\left[E\left[\max\{R, V(T^*)\}\right] - \sum_{i=1}^{T^*} s_{\sigma_{T^*}(i)}\right]\right|_{R=0} \geq \mu > 0.$$

Moreover, it holds that $\lim_{R\to\infty} E\left[\max\{R, V(T^*)\}\right] - \sum_{i=1}^{T^*} s_{\sigma_{k^*}(i)} - R = 0$. Thus, for all $h \leq \overline{h}$ an R^* exists that solves equation (11). Uniqueness follows from the consumer believing that if there is more than one outside option consistent with some h the lowest outside option is the correct one. With this, R^* is strictly decreasing in h. **Q.E.D.**

The above Proposition allows to establish that the number of attributes learned in equilibrium is hump-shaped in the search costs h. This implies that the consumer learns less about a firm's product both when costs of finding a new match are rather small and when they are rather high. Interestingly, this is in stark contrast to the way learning costs affect the learning decision which decreases as learning costs increase.

Corollary 5 *There exists a h_μ such that the equilibrium learning decision T^* is increasing in $h < h_\mu$, and it is decreasing in $h > h_\mu$.*

Proof. I obtain the existence of h_μ from the uniqueness result in Proposition 4. Moreover we have that the equilibrium outside option R^* is decreasing in h. This implies with Corollary 1 that the equilibrium number of learned attributes T^* must be increasing in $h < h_\mu$, and it must be decreasing in $h > h_\mu$. **Q.E.D.**

Continuous Learning with Differences in Dispersion

Consider a version of the problem where the receiver learns simultaneously about an offering with an infinite number of attributes. In particular, I impose that there is a unit mass of attributes $v(i)$ indexed by $0 \leq i \leq 1$ that differ in their rotation. The rotation function $\theta(i) > 0$ is strictly decreasing and continuous in i. Analogous to Section 6, the posterior dispersion of $V(T, \sigma)$ is given by

$$\Theta(T, \sigma) = \int_0^T \theta(\sigma(i))di.$$

I stipulate that $V(T, \sigma)$ is unimodal and symmetric around mean 0 for all T and all orders σ.[16] Learning a mass of T attributes costs Ts and thus the (net) value of learning a mass of T attributes with an order σ is given by

$$W(\Theta(T, \sigma)) - Ts$$

where $W(\Theta(T, \sigma)) \equiv E[\max\{R, V(T, \sigma)\}]$ is the gross value of information which is only affected by the posterior dispersion Θ. Clearly, the gross value of information is increasing in the total dispersion. A dispersion Θ' is called *implementable* if an order σ exists such that $\Theta' = \Theta(T^*, \sigma)$. In order to streamline the analysis impose that $W(\Theta)$ is continuous and strictly increasing in Θ with $W(\Theta(1, \sigma)) < \infty$. Analogous to above, the receiver chooses T^* in order to maximize $W(\Theta(T, \sigma)) - Ts$ and with the continuity of $W(\cdot)$ an optimum T^* always exists. I can now provide a characterization of optimal orders with dispersion heterogeneity similar to the one provided by Lemma 3 for the case of cost heterogeneity with a finite number of attributes.

Lemma 9 *Suppose the receiver learns simultaneously about an offering with an infinite number of attributes that differ in dispersions, and suppose that the order σ results in a posterior dispersion of $\overline{\Theta} = \Theta(T^*, \sigma)$. Then the order $\sigma_{\overline{\Theta}}$ with*

$$\sigma_{\overline{\Theta}} = \begin{cases} k - i & \text{for } i \leq k \\ 1 - i + k & \text{for } i > k \end{cases}$$

with k being defined by $\overline{\Theta} = \int_0^k \theta(i)di$ also results with optimal receiver learning in a posterior dispersion of $\overline{\Theta}$: $\Theta(T^*(\sigma), \sigma) = \Theta(T^*(\sigma_{\overline{\Theta}}), \sigma_{\overline{\Theta}})$.

[16] One micro-foundation for this is given by the following. Suppose the value of the offering is given by a Brownian motion that has a time-dependent instantaneous variance $\theta(\sigma(i))$ and is stopped at time $T \leq 1$. Then, $V(T, \sigma)$ is normally distributed with variance $\int_0^T \theta(\sigma(i))di$. Note that the cdf of a normal distribution rotates in its variance. See Branco et al. (2012) for an application.

Proof. By definition, it holds that $W(\Theta(T^*(\sigma),\sigma)) = W(\Theta(k,\sigma_{\overline{\Theta}}))$ and for any T' there exists a T'' with $W(\Theta(T',\sigma)) = W(\Theta(T'',\sigma_{\overline{\Theta}}))$. When $T' < T^*(\sigma)$, it must hold for the corresponding T'' that $T^*(\sigma) - T' > k - T''$ since for $\Delta > 0$ (but not too large) it holds that

$$\int_{T^*(\sigma)-\Delta}^{T^*(\sigma)} \theta(\sigma(i))di \leq \int_{k-\Delta}^{k} \theta(\sigma_{\overline{\Theta}}(i))di,$$

and, by definition of $\sigma_{\overline{\Theta}}$, we have

$$\int_{0}^{T^*(\sigma)} \theta(\sigma(i))di = \int_{0}^{k} \theta(\sigma_{\overline{\Theta}}(i))di.$$

Thus, we have for $T' < T^*(\sigma)$ and the corresponding $T'' < k$ that

$$\begin{aligned} 0 &\leq W(\Theta(T^*(\sigma),\sigma)) - W(\Theta(T',\sigma)) - s(T^*(\sigma) - T') \\ &\leq W(\Theta(k,\sigma_{\overline{\Theta}})) - W(\Theta(T'',\sigma_{\overline{\Theta}})) - s(k - T''). \end{aligned}$$

Therefore, the receiver still prefers to learn at least until the same mass of attributes is learned. By the same arguments it follows that learning a mass of attributes that results in a higher dispersion than $\overline{\Theta} = \Theta(T^*(\sigma),\sigma)$ when the order is given by $\sigma_{\overline{\Theta}}$ results in strictly higher costs than with σ, so that the claim follows. **Q.E.D.**

Denote again with Θ^* the equilibrium rotation that results from the sender picking the optimal order while anticipating the receiver's best response. I close this Section with a Proposition that describes the equilibrium and shows how Θ^* changes when either the dispersion of attributes increases or when the concentration of dispersions increases, that is, attributes with a high dispersion become more dispersed while attributes with a low dispersion become less dispersed. For the sake of brevity, I stipulate that a sender optimal order always exists. The intuition for this result is the same as that of Corollary 2. The main difference is that in the current case incentives to learn are set rather by changes in the gross marginal benefits of learning rather than the costs of information acquisition.

Proposition 5 *Suppose the receiver learns simultaneously about an offer with an infinite number of attributes that differ in dispersions, and the sender determines the order in which the receiver learns. Then, the equilibrium is characterized as follows. When $R < 0$ ($R > 0$), the equilibrium dispersion Θ^* equals the smallest (largest) implementable dispersion. The* *sender chooses the order* σ_{Θ^*}*, and the receiver learns a mass of attributes so that the resulting dispersion is* Θ^**. When some* $\theta(i)$ *increase by* $\Delta\theta > 0$*, the equilibrium dispersion* Θ^* *increases. When the dispersion function* $\theta(\cdot)$ *shifts to a more concentrated dispersion function* $\widetilde{\theta}(\cdot)$*, so that there is a* $0 < j < 1$ *such that* $\widetilde{\theta}(i) > \theta(i)$ *for* $i < j$ *and* $\widetilde{\theta}(i) < \theta(i)$ *for* $i > j$ *but still* $\int_0^1 \widetilde{\theta}(i)di = \int_0^1 \theta(i)di$*, the equilibrium dispersion* Θ^* *decreases (increases) when* $R < 0$ *(*$R > 0$*).*

Proof. The first claim is a direct consequence of the rotation property and Lemma 9.

The first comparative result of the proposition follows immediately from the fact that an increase in $\theta(i)$ for some i increases the benefits of learning for any order.

For the second comparative result, I show that a shift from $\theta(\cdot)$ to $\widetilde{\theta}(\cdot)$ increases the incentives for all orders σ_Θ to learn a mass of attributes that, given σ_Θ, results in a dispersion of Θ. For this define $\xi_\Theta(\Theta')$ as the mass of attribute that the receiver must learn to obtain a dispersion Θ' when the order is σ_Θ:

$$\Theta' = \int_0^{\xi_\Theta(\Theta')} \theta(\sigma_\Theta(i))di.$$

Formally, I show that

$$W(\Theta) - \xi_\Theta(\Theta)s - [W(\Theta') - \xi_\Theta(\Theta')s]$$

increases as $\theta(\cdot)$ shifts to $\widetilde{\theta}(\cdot)$ for $\Theta' \neq \Theta$. Note first, that for given Θ and Θ' the terms $W(\Theta)$, $W(\Theta')$, and s are constant. Thus, it remains to show that $\xi_\Theta(\Theta) - \xi_\Theta(\Theta')$ decreases. Denote with $\widetilde{\xi}_\Theta(\Theta')$ the respective mass of attributes when the dispersion function is given by $\widetilde{\theta}(\cdot)$. Consider first the case $\Theta' < \Theta$. Then, by definition it holds that

$$\Theta - \Theta' = \int_{\xi_\Theta(\Theta')}^{\xi_\Theta(\Theta)} \theta(\sigma_\Theta(i))di = \int_0^{\xi_\Theta(\Theta)-\xi_\Theta(\Theta')} \theta(i)di$$

and similarly

$$\Theta - \Theta' = \int_0^{\widetilde{\xi}_\Theta(\Theta)-\widetilde{\xi}_\Theta(\Theta')} \widetilde{\theta}(i)di.$$

Since $\widetilde{\theta}(i) > \theta(i)$ for $i < j$ and $\widetilde{\theta}(i) < \theta(i)$ for $i > j$ with $\int_0^1 \widetilde{\theta}(i)di = \int_0^1 \theta(i)di$ and $\widetilde{\theta}(i), \theta(i) > 0$, it follows that $\xi_\Theta(\Theta) - \xi_\Theta(\Theta') > \widetilde{\xi}_\Theta(\Theta) - \widetilde{\xi}_\Theta(\Theta')$ and so the claim follows. When $\Theta' > \Theta$ it holds by following the same steps as above that

$$\Theta' - \Theta = \int_{1-[\xi_\Theta(\Theta')-\xi_\Theta(\Theta)]}^{1} \theta(i)di$$

and

$$\Theta' - \Theta = \int_{1-\left[\widetilde{\xi}_\Theta(\Theta') - \widetilde{\xi}_\Theta(\Theta)\right]}^{1} \widetilde{\theta}(i) di$$

from which it immediately follows that $\xi_\Theta(\Theta) - \xi_\Theta(\Theta') > \widetilde{\xi}_\Theta(\Theta) - \widetilde{\xi}_\Theta(\Theta')$. **Q.E.D.**