The Signaling Value of Simplicity

Author(s):
Dolecek, Roman; Dubus, Antoine; Sahuguet, Nicolas; Waelbroeck, Patrick

Publication date:
2025-01-06

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-b-000713838

The Signaling Value of Simplicity*

Roman Dolecek,† Antoine Dubus,‡

Nicolas Sahuguet,§ and Patrick Waelbroeck¶

January 6, 2025

Abstract

Complex signals are often misunderstood by their receivers, leading to failures of information transmission in signaling games. In this study, we introduce novel methods to systematically assess the complexity of a wide range of signals, using the concepts of *theoretical* and *analogical* complexity. Theoretical complexity measures the amount of information a signal contains when compressed, while analogical complexity captures its social relevance. To illustrate our approach, we analyze data from the sale of license plates in Switzerland, where specific numbers serve as indicators of social status. By evaluating both the theoretical and analogical complexity of all license plates, we show that plates with simpler designs tend to sell for higher prices. Our findings suggest that simplicity is a key factor of a signal's perceived value, offering broader implications for signaling theory and decision-making in contexts involving complex reasoning.

Keywords: Complexity; Signaling; Conspicuous consumption.
JEL: D12; Z13.

*We would like to thank Laura Abrardi, Rava Azeredo da Silveira, Nora Bardelli, Jean-Louis Dessales, Patrick Legros, Christian Stettler, as well as participants to the 2023 EARIE conference and the the CER-ETH seminar for useful comments and suggestions. We also thank the road traffic office of the Canton of Zürich, Switzerland, for giving us access to their auction data.

†Department of Management, Technology and Economics, ETH Zürich Leonhardstrasse 21 Switzerland – 8092 Zürich, Switzerland; rdolecek@ethz.ch.

‡Department of Management, Technology and Economics, ETH Zürich Leonhardstrasse 21 Switzerland – 8092 Zürich, Switzerland; antoine1dubus@gmail.com.

§Department of Applied Economics, HEC Montreal, 3000 St Catherine Road, Montreal, QCH3T2A7, Canada; nicolas.sahuguet@hec.ca.

¶i3, Telecom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, France; patrick.waelbroeck@telecom-paris.fr.

1 Introduction

Signals are fundamental to human interactions, influencing a wide range of domains such as education, consumption, labor markets, financial investments, or innovation (Riley, 2001). However, signals are often complex, costly to process mentally, or require senders and receivers to share common knowledge. While canonical signaling models assume that receivers can perfectly understand a signal and accurately infer the sender's type, many real-world economic situations deviate from these idealized assumptions.

Effective communication often requires significant efforts: the sender must distill complex information into a clear and concise message, while the receiver must invest energy to comprehend it (Dewatripont and Tirole, 2005). In the context of signaling, this implies that simpler signals hold value as they can reach – and be understood by – a broader audience.

Consider the job market signaling model of Spence (1978, 2002), in which applicants demonstrate their ability by obtaining an education that requires publicly observable effort. However, education systems are often highly complex, with thousands of universities offering degrees that vary widely in quality. In many cases, employers may struggle to recognize high-quality education programs as understanding the value of a degree from an unfamiliar institution requires significant time and effort. This additional cost of interpreting a signal can discourage employers from fully assessing an applicant's qualifications. For example, ENS (École Normale Supérieure), one of the most prestigious institutions in France, is relatively unknown in the US.[1] As a result, a degree from ENS carries limited signaling value in the American job market. In contrast, a degree from Harvard University is widely recognized, allowing applicants to use it effectively as a signal of their qualifications.

Similarly, consider a status signaling game through conspicuous consumption (Veblen, 1899; Bernheim, 1994). For individuals seeking to signal their wealth, merely spending a large sum on a publicly visible good is not enough to ensure that others perceive its true value. Observers must also recognize that the item was purchased at a high price. Luxury fashion brands must ensure that their products serve as effective signals for status. To achieve this, they incorporate easily recognizable features that unmistakably identify the brand – such as Louboutin's signature red sole or Louis Vuitton's iconic Damier canvas pattern. These distinctive elements reduce the cognitive effort required for observers to associate the product with its brand, thereby minimizing the risk of failing to recognize its status value.

Beyond luxury fashion, the importance of simple signals in conspicuous consumption is evident in various contexts. Consider, for instance, expensive wines, which can sell for thousands of dollars per bottle at auctions. A host aiming to display its social status might

[1] https://voices.uchicago.edu/euchicago/why-have-you-heard-of-oxford-but-not-ecole-normale-superieure/

serve an expensive wine without explicitly mentioning its price, as doing so is often socially frowned upon. However, the complexity of factors such as grape variety, vintage, terroir, and vineyard makes it nearly impossible for a non-expert to accurately assess a wine's quality or infer its price based on these attributes alone (Ali et al., 2008). As a result, such signals are difficult to interpret and inefficient for signaling status outside specialized circles. This complexity helps explain the widespread adoption of the Parker grade, a system that assigns each wine a score between 50 and 100, with higher scores typically indicating higher quality and price. For buyers, this grading system simplifies the decision-making process, allowing them to navigate the diverse range of options more easily. Additionally, the score can serve as a clear status signal: by disclosing a wine's Parker grade, a host can communicate its quality and value, even to those without expertise. In this way, grading systems offer a practical simplification of complex signals, enhancing the efficiency of signaling games.

While signal complexity is a fundamental aspect of any signaling game, assessing the value of simple signals raises three main challenges. First, defining and quantifying signal complexity in specific economic contexts is inherently difficult. For instance, what constitutes a valid measure of complexity for an education degree? Or how can we determine whether a consumption good serves as a simpler signal of status compared to another? Second, it must be shown from a theoretical standpoint that simplicity itself can carry an economic value. Third, from an empirical perspective, identifying a suitable context to measure the economic value of simplicity is challenging, as it requires isolating the signaling value of a good. In the case of conspicuous consumption, for instance, status goods often possess functional value, such as luxury cars being of superior quality which partially explains their high price. Disentangling the signaling value from these functional attributes is essential to understanding the economic role of simplicity.

In this article, we develop a model of status signaling in which consumers purchase conspicuous goods that vary in complexity (or, conversely, simplicity). We propose several measures to quantify the complexity of the signals these goods convey to observers. Using these measures, we demonstrate that auction participants are willing to pay a premium for simpler signals, which are more easily understood by a broader audience, thereby enhancing their conspicuous effect. To empirically estimate the signaling value of simplicity, we use data on the sale of license plates in Switzerland. Individuals pay remarkably high prices for specific plate numbers to signal their wealth. This data set provides an ideal framework to estimate the value of simplicity: a specific plate number has no intrinsic value beyond signaling status, as there are no inherent qualities or aesthetics associated with different license plates.[2]

[2]In some cases, simple numbers may have minor practical benefits, such as being easier to remember for administrative purposes. However, this utility is negligible compared to their potential for status signaling.

Additionally, car owners who choose not to pay a high price can obtain a plate number for free. In the canton of Zürich, plate owners are not permitted to resell their plates, eliminating the possibility of future monetary gains from the purchase.[3] Our analysis reveals that a key determinant of a plate's value is its simplicity.

We introduce a novel methodology for quantifying signal complexity, based on two distinct concepts: theoretical and analogical complexity. Theoretical complexity measures the information content of a signal, while analogical complexity accounts for the specificity of a social or cultural context, including factors like reputation effects.[4]

The concept of theoretical complexity is grounded in information theory and refers to the amount of information embedded within a signal, which is classically measured using Kolmogorov complexity (k-complexity) (Kolmogorov, 1963).[5] For a given language or algorithm, the k-complexity of a signal is defined by the length of its shortest description. In the context of license plates, this implies that shorter numbers have lower k-complexity. In Zürich, license plates consist only of digits (along the prefix ZH found on all plates), so that, for instance, 100 is simpler than 986451. Moreover, mental operations of compressions using specific patterns allow an observer to lower the complexity of a signal (see e.g. Dehaene et al. (2022)). For instance, palindromes and repeated digits are simple to process mentally, reducing the k-complexity of a sequence. Beyond license plates, our earlier example of status signaling through wine consumption illustrates the value of a simple grade compared to a more complex set of information.

However, this concept does not account for the fact that the complexity of a signal may vary based on social or cultural contexts. For instance, number 911 is not only relatively short but also serves as an emergency number in the US and has gained prominence through popular movies, songs and other cultural references. We characterize this cultural significance using the concept of analogical simplicity, which helps explain why 911 is easier to remember than 922, despite their comparable theoretical simplicity. Systematically measuring the analogical simplicity of a signal is challenging, as perceptions of simplicity can differ widely across social and cultural groups. We address this challenge by constructing an index based on the frequency of queries in the Google search engine. Numbers with higher search results are likely to appear in diverse contexts, such as products, movie titles, or phone numbers. We use this resulting index as a proxy for the analogical simplicity of the number.[6] The index

[3] This rules out speculative motives for plate acquisition, an effect that takes place for instance in the case of collectible stamps (Dimson and Spaenjers, 2011).

[4] Kurtz-David et al. (2024) make a similar distinction in their analysis of agents performing different tasks, which may involve social or arithmetic complexity.

[5] This notion is related to the amount of information contained in a signal measured by Shannon entropy (Shannon, 1948), and even as asymptotically equivalent when it comes to the expected k-complexity.

[6] This technique was introduced by Cilibrasi and Vitanyi (2007) in their work on Google similarity distance,

provides a measure of the social significance of each number. This measure of analogical complexity can be easily generalized to any type of product and many types of information sets. For instance, the relative reputation of universities can be assessed with this measure, and we can easily verify that Harvard University generates many more queries than other, less well-known universities.

Overall, our empirical setting of license plate auctions offers the advantage of incorporating both measures of complexity, enabling us to evaluate and compare their relative benefits.

Our contribution is twofold. On the one hand, we provide the first estimation of the signaling value of simplicity. We therefore contribute to the literature that usually considers theoretical approaches or lab experiments by using real-world market data. On the other hand, using our empirical study we validate the conceptual framework that we introduce based on the distinct notions of analogical and theoretical complexity. We show that each notion has a significant impact on the market value of signals. Finally, we provide benchmark tools that can be used in further research to characterize the complexity of a wide range of goods, information sets, and economic decisions.

The remainder of this paper is organized as follows. We review the literature on complexity, conspicuous consumption, and signaling in auctions in Section 2. Section 3 describes the data and provides the context of the signaling game that we use to illustrate our theory. We introduce the concepts of theoretical and analogical complexity in Sections 4 and 5. In Section 6, we develop a simple theoretical auction model in which agents can use the product for signaling, with an efficiency depending on the complexity of the good. Section 7 presents our empirical analysis, and we consider broad implications in Section 8. Finally, Section 9 concludes.

2 Literature

Measures of complexity.

The economic literature has employed various measures to analyze the impact of complexity of economic decisions, all of which relate to the amount of information an agent must process (see Gell-Mann and Lloyd (1996) and Lloyd (2001) for classifications of the different measures of complexity used in the literature). More broadly, these concepts are grounded in the general definition of Kolmogorov complexity found in theoretical computer science literature, which serves as the foundation for the first measure of complexity we use in our analysis. This measure is closely related to the minimum description length criterion introduced by

and has since been widely adopted by computer scientists. (We would like to thank Jean-Louis Dessalles for pointing out this reference.)

Rissanen (1978) and is commonly applied in models of machine learning (Barron et al., 1998; Hansen and Yu, 2001).[7] Additionally, k-complexity is frequently used in the cognitive sciences literature (Soler-Toscano et al., 2014; Amalric et al., 2017; Murena et al., 2020; Planton et al., 2021).

Our measures of theoretical and analogical complexity encompass the different approaches discussed below. By facilitating the assessment of decision-making complexity and the intricacy of economic environments, our methodology can improve models in these areas of research by enabling nuanced characterizations of the complexity inherent in signals, information sets, and economic goods across various economic contexts.

Decision making. We contribute to the literature on decision-making in complex environments (Woodford, 2020; Gabaix and Graeber, 2023), which has long recognized the significance of complexity in decision-making processes (Simon, 1955). In this literature, complexity is typically measured by the amount of information an agent must process, such as the number of options in a choice set or the support of a lottery.[8]

In repeated games with multiple states, Oprea (2020) and Banovetz and Oprea (2023) characterize the complexity of a decision based on the number of states it requires, in line with the automata literature.

Enke and Zimmermann (2019) and Enke (2020) define the complexity of the decision faced by agents using the number of messages that they receive. In the case of chess players, Salant and Spenkuch (2022) use an adapted version of the Kolmogorov complexity of a move to show that complex moves are chosen less often than simpler ones.

In Mador et al. (2000), Sonsino et al. (2002), Iyengar and Kamenica (2010), Bernheim and Sprenger (2020), Fudenberg and Puri (2022), Oprea (2024), and Puri (2024) participants to a lab experiment face lotteries with different numbers of outcomes and the complexity of a lottery depends on its distribution.

Sims (2003), Woodford (2001) and Azeredo da Silveira et al. (2024) analyze forecasting problems under which agents make decisions according to past realizations of the states of the world, considering agents that are constrained respectively on the process of information flows, on access to information, and on memory.

Legal systems. In Schuck (1992), Abeler and Jäger (2015), and Bennani and Neuenkirch (2023), the complexity of a tax system is characterized by the number of rules it encompasses, while Slemrod (2005) focuses on the number of lines in tax forms. Similarly, Colliard and

[7] Chater and Vitányi (2003) review the literature on simplicity in cognitive sciences.

[8] A recent theoretical literature also analyzes the role of complexity in communication (Jäger et al., 2011; Sobel, 2012; Hertel and Smith, 2013).

Georg (2023) employs a comparable measure to assess regulatory complexity. Additionally, Ash et al. (2024) and Foarta and Morelli (2022) consider more sophisticated factors within legal systems and contracts, taking into account their completeness and the conditions imposed on contingencies.

Complex organization. In the case of effort provision in a working environment Abeler et al. (ming) define the complexity of an incentive scheme using the number of contingencies. A similar notion is used by Jakobsen (2020) to define the complexity of a contract.[9]

Product and economic complexity. Hidalgo (2009), Hidalgo and Hausmann (2009) and Hartmann et al. (2017) use the size of the set of capabilities located in the country to characterize its economic complexity.

Carvalho and Silverman (2024) define the complexity of a financial investment by the number of options available. Célérier and Vallée (2017) use three measures to characterize the complexity of financial products: number of payoffs, number of products, and description length. Similarly, Ghent et al. (2019) characterize the complexity of a deal using different measures of the number of contingencies. Waelbroeck (2003) and Thakor and Merton (2023) use similar notions to characterize the complexity of products.

K-complexity for information transmission.

In the case of signals, the computer science literature defines k-complexity as the minimum length of code required to describe information (in a universal Turing machine, see e.g. Li et al. (2008)). While this measure is intuitive, a major challenge arises from the difficulty to characterize the minimum description length for any type of signal.[10] Indeed, there is usually the possibility for a receiver of the message to compress its length while keeping its informational content. For instance, abbreviations such as "e.g." instead "for instance" allow to shorten greatly the total length of a text. This issue is well known in the computer science literature, which has developed techniques of compression that approximate the k-complexity of large information sets (Ziv and Lempel, 1977).[11] In the case of plate numbers, we thus

[9]Cremer et al. (2007) analyze communication within a firm and between firms when the environment can be more or less complex. Their definition of complexity corresponds to the predictability of the environment, and is equivalent to first-order stochastic dominance.

[10]There exist efficient approximations of the k-complexity of short sequences of digits with an alphabet of maximal size equal to 9 (Soler-Toscano et al., 2014). We cannot use such programs as the size of the alphabet for license plates is equal to 10 (all numbers between 0 and 9). Other approximations include the Python koolmogorov library. We focus on the methodology described below as it allows to explain the mechanisms behind the compression operations.

[11]For instance, a classical estimate the k-complexity of a text is its number of bytes after compression by a computer program such as gzip (Juola, 2008; Szmrecsanyi, 2016).

need to characterize the minimum description length of any signals, accounting for a possible compression.

The k-complexity of a signal also determines the ability of a receiver to process it, as discussed by Kemp (2012) among others. In the case of an agent, an operation of mental compression of information induces a cognitive compression cost (Dehaene et al., 2022). For instance, the number 50653 can be compressed by writing it as 37^3, but this operation induces a high cognitive cost, and accounting for this cost, the raw description of the number may be simpler to process cognitively for an agent.[12]

Overall, using k-complexity to model the cognitive costs of an agent engaging into a mental operation of compression of a signal is in line with the language of thought hypothesis according to which humans can compress regular sequences using mental operations (Fodor, 1975). Hence, using this approach, when an agent receives a signal it will be able to process it only if the uncompressed complexity is low enough, or if the k-complexity accounting for the compression cost is low enough. Otherwise, the agent will not understand the signal.

Shannon entropy.

Our objective is to provide a theoretical framework that explains the value of license plates based on their simplicity. While we focus on Kolmogorov complexity, other approaches may be relevant. For instance, the literature has identified important equivalences between expected Kolmogorov complexity and Shannon entropy (Leung-Yan-Cheong and Cover, 1978; Hammer et al., 2000; Grunwald and Vitányi, 2004). In our context, we could characterize the Shannon entropy of license plates using an algorithm similar to the one defined for our measure of theoretical complexity.

Shannon entropy has been widely employed in the rational inattention literature, where agents select the information flows they use to make decisions (Sims, 2003; Maćkowiak et al., 2023). However, our framework differs from rational inattention models; the problem faced by an agent is not about selecting information, but rather assessing the value of a plate based on its complexity. Therefore, the notion of k-complexity is especially well suited to our analysis.

Additionally, k-complexity is relevant for agents evaluating whether a specific displayed number was chosen randomly or selected by the person showing it. This distinction is important in our setting, as those who do not purchase a specific plate are assigned a random plate number by the relevant authority.

From a conceptual standpoint, focusing on k-complexity allows us to characterize a relative measure for which a specific programming language running on a universal Turing machine

[12]The role of cognitive costs in decision making and how they can explain deviations from optimal behaviors is also the topic of a recent economic literature (Charness and Levin, 2005, 2009; Enke et al., ming).

has only a limited impact according to the invariance theorem (Calude, 2002; Gauvrit et al., 2016). On the contrary, entropy-based compression algorithms depend on the language choice and may therefore not be robust enough to measure complexity or randomness (Zenil et al., 2017).

Complex reasoning theory.

Our methodological distinction draws interesting connections with the literature on complex reasoning, from which we derive our terminology for analogical and theoretical complexity. According to Gayer and Gilboa (2014), agents tend to rely on theoretical reasoning when the complexity of a decision is low, whereas they resort to analogical thinking when the rules become more complex. Moreover, in scenarios requiring coordination among a large number of agents, theoretical reasoning is likely to prevail. In our framework, analogies are captured by the notion of cultural significance, which we measure using our Google Frequency Index. Conversely, theoretical thinking corresponds to the computation of simplicity, using the properties of a number as represented by our measure of theoretical complexity.

Conspicuous consumption.

We apply the measures of complexity we have developed to analyze an empirical model in which consumers aim to signal their wealth through consumption, therefore engaging with the literature on conspicuous consumption. Our approach is related to the model of Ireland (1994), in which agents can allocate part of their consumption of a numeraire good toward conspicuous consumption. We adapt this model into a context where agents compete to acquire a conspicuous good through an auction mechanism. While existing literature has explored the benefits of status signaling, particularly through social interactions with wealthier individuals (Bagwell and Bernheim, 1996), our focus is on self-image motives, which fit well with our empirical application as license plate owners do not directly interact with observers of their plates.

Signaling in auctions.

We support our empirical evidence using a theoretical model of signaling in auctions. Hence, we also relate to the literature on signaling by auction participants to an outside observer (and not to other bidders as in Hörner and Sahuguet (2007), who analyze jump bidding and bluffing as signaling strategies in dynamic auctions).

There is limited literature on this topic, starting with Haile (2003) and Goeree (2003), who consider an auction followed by competition among bidders, for instance if the good can be resold. Bids are used by auction participants to signal the value of the good for sale.

In a similar spirit, Giovannoni and Makris (2014) consider reputational concerns among auction participants, who care about the belief of an observer regarding their type. Bos and Truyts (2021) analyze a related question when only the final bid and the winner's identity are observable.

Different from these settings, bids are not observable in our model, but only the identity of the winner and the complexity of the object for sale. Simpler signals can be understood by more observers, and we show that the value of the bid increases with simplicity. Moreover, we explicitly model the disutility from paying a higher bid through a consumption function, providing a microfoundation for conspicuous consumption. Intuitively, we show that the wealthiest bidder wins the auction, as their marginal disutility from winning is lower than that of other bidders.

Marketing of luxury goods.

Finally, we contribute to the marketing literature that analyzes branding and pricing strategies for luxury goods. In particular, Han et al. (2010) examines how brand prominence interacts with consumers' desire to signal their wealth to either a large audience or to a selected group from the same social circle. Yuan et al. (2022) estimates the price premium of luxury goods related to their signaling value. We discuss the implications of our results for marketing in Section 8.

3 Background and Data Description

License plates are sold in more than thirty countries, including high-profile sales such as in the United Arab Emirates where a plate fetched $14 million, and in the UK, where certain plates have sold for over £500,000. Other countries with active license plate markets include the Czech Republic, Norway, Canada, China, Taiwan, Hong Kong, Malaysia, and the US—where even Milton Friedman famously displayed his quantity equation on his California license plate.[13]

In this paper, we focus on the license plate market in the Swiss canton of Zürich, where the sale of license plates generates over 5 million CHF annually. In Switzerland, it is common for car owners to pay large sums for specific license plate numbers, providing a steady source of income for local governments. A Swiss license plate consists of a two-letter code indicating the canton and a number with up to six digits. Figure 1 displays license plate 831 in canton Zürich, which was auctioned for 30,000 CHF.

[13]References for these different examples are given in Appendix A.1.

Figure 1: License plate number 831 in canton Zürich.

For our analysis, we use a dataset provided by Zürich's road traffic office. The dataset covers the period from August 2017 to January 2023 and includes all 6,659 auctions conducted by the office. Each record contains publicly observable information from the second-price auctions, including final sale prices and plate attributes.

3.1 Descriptive Statistics

Plate frequency. Table 1 summarizes the frequency of the most common plate attributes in our dataset as well as their share in $\mathbb{N}$ (to verify that our sample is representative). Some features do not apply to all plate numbers by construction. For instance, three-digit plates cannot be multiples of 1000. This is also the case for attributes "double" and "repeated pair". Attribute "double" corresponds to two numbers repeated twice, such as 4422. Similarly, attribute "repeated pair" corresponds to repetitions of the form 4242. By definition, these attributes only apply to 4-digit plates, and not to other plate lengths. We denote these cases using symbol "–". Also note that a given plate number can hold multiple features. For example, number 8888 is a palindrome and consists of one unique digit.

Table 1: Summary of plate attributes

	Share of characteristics in Sample vs in $\mathbb{N}$						
	palindrome	1 unique digit	2 unique digits	repeated pair	doubles	multiple of 100	multiple of 1000
3 (n=194)	10.825%	1.031%	30.928%	-	-	1.546%	-
share in $\mathbb{N}$	10%	1%	27%	-	-	1%	-
4 (n=273)	1.832%	0.733%	6.960%	1.832%	1.832%	0.733%	0.000%
share in $\mathbb{N}$	1%	0.1%	6.3%	1%	1%	1%	0.1%
5 (n=6,192)	0.581%	0.000%	0.727%	-	-	0.630%	0.048%
share in $\mathbb{N}$	1%	0.01%	2.35%	-	-	1%	0.1%
Total (n=6,659)	0.931%	0.060%	1.862%	0.075%	0.075%	0.661%	0.045%

The most common feature is two unique digits (e.g., 73773), making up a large share of the shorter number plates. It is followed by palindromes and multiples of 100. In particular, these features are present in 63.3%, 31.6% and 22.4% of all plates with at least one of the listed attributes.

Price. The price of a plate is determined in a second-price auction.[14] Table 2 depicts the distribution of prices for plate numbers with three, four, and five digits.

Table 2: Distribution of prices by number of digits on a plate

	Value of a Plate								
Number of digits	Min	P5	P10	P50	P90	P95	Max	Mean	Sd
3 (n=194)	16,000	18,800	20,000	26,000	45,200	60,000	226,000	32,282	25,123
4 (n=273)	6,000	7,600	8,100	10,800	20,000	22,800	55,000	12,688	5,980
5 (n=6,192)	700	1,000	1,040	1,530	3,000	3,700	27,000	1,840	1,092
Total (n=6,659)	700	1,000	1,050	1,580	4,150	11,000	226,000	3,172	7,136

The average price is 3,172 CHF with a standard deviation of 7,136. The distribution is highly skewed with the 95th percentile at 1,100 CHF. The mean of log prices is equal to 7,579 and the standard deviation is 0.732.

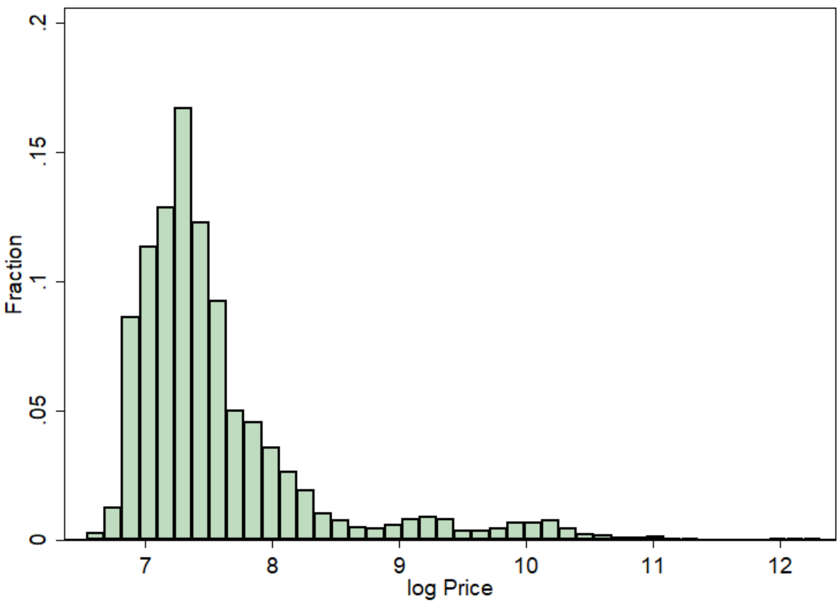

Figure 2: Histogram of log of license plate prices.

The Most Valuable Plates. Table 3 reports the price of the most valuable plates.

[14]It is important to note that the initial sample includes some six-digit plates. Because not all six-digit plates are sold via these auctions – most are distributed for free upon request – there is a selection bias for these plates, as only those expected to have the highest value are auctioned. To avoid this bias in our analysis, we focus exclusively on license plates with up to five digits, as all such plates are auctioned by the authority.

Number	Price in CHF
100	226000
888	194000
911	173600
987	152400
556	76800
812	72000
551	60800
555	60400
621	60000
707	60000

Table 3: Plate numbers ranked by prices.

Some common patterns emerge among these plates. First, they tend to be simple, often comprising only three digits. Moreover, they frequently feature either repeated numbers, palindromes, or multiples of 100. Interestingly, lower numbers are not necessarily the most valuable. For example, a culturally significant number such as 911 reaches the third-highest price. Other notable plates include 1965 auctioned for 40,000 CHF, and 5454, which sold for 35,000 CHF.

To explain the characteristics of the most valuable plates, we introduce the concepts of theoretical and analogical simplicity in the following sections.

4 Theoretical Complexity

In this section, we introduce a method to quantify the complexity of license plates by constructing a language specifically designed to compress numerical sequences.[15] The use of language to represent cognitive tasks is well-established in cognitive science, dating back to the influential work of Fodor (1975). We apply this concept to describe the k-complexity of short sequences of digits, such as license plates. Specifically, we simplify plate numbers using prefixes that correspond to specific operations, a classical method to characterize the k-complexity of finite sequences (Lempel and Ziv, 1976). This approach is also common in computer science, where tailored languages are built to assess simplicity in various contexts.[16]

[15] Several programs approximate the k-complexity of many information sets (Lempel and Ziv, 1976; Soler-Toscano et al., 2014; Aksentijevic et al., 2020), yet existing methods do not allow for characterizing the k-complexity of short sequences using an alphabet of size 10, which applies to license plates.

[16] Amalric et al. (2017) analyze the complexity of geometrical figures; Murena et al. (2020) measure analogies between words; and Planton et al. (2021) focus on general binary sequences.

A key assumption of our model is that when observers see a license plate, they engage in one of the mental processing tasks defined in our language to compress the information contained in the signal.[17] Our empirical results support this approach. The language we construct showcases the main compression operations an individual might use when seeing a license plate. Below, we describe the language, which is specifically designed for the context of numbers between three and five digits in the decimal system. For each number, we compute its simplicity within this language (referred to as its theoretical simplicity) by measuring the length of the code used for compression. We also examine how various features of a number influence its perceived simplicity.

Description of the language.

The length of a number in decimal code corresponds to its uncompressed complexity. The different prefixes correspond to characteristics of numbers that allow for compression:

- A: baseline complexity without simplification.
- B: palindrome (e.g., 16861).
- C: pairs (e.g., 4242).
- D: doubles (e.g., 4422).
- E: composed of a unique digit (e.g., 555).
- F: multiple of 100.
- G: multiple of 1000.
- H: multiple of 10,000.

We apply this language to determine the simplicity of each plate number in our dataset. For every license plate, we evaluate the k-complexity using all relevant simplification operators, and the minimum resulting value corresponds to the theoretical simplicity of the number.

The selection of these simplification operations is motivated by a well-established cognitive science literature, which explores individuals' sensitivity to specific patterns within sequences. Our first step involves calculating the baseline k-complexity without simplification (A). This is important, as assessing the cardinality of a set is recognized as one of the most basic human

[17] Research supports this claim by showing that perceptual organization is based on simplicity (Chater, 1996; Chater and Vitányi, 2003).

cognitive operations (Wynn, 1992; Piantadosi et al., 2012). Accordingly, the number of digits present in a license plate is a core determinant of its complexity.

For operations B, C, D, and E, the sensitivity of individuals to repetitions, symmetry, and "cyclical" symmetry (alternating patterns) captured by the prefixes pairs and doubles has been established by Glanzer and Clark (1963) and widely used since then (e.g., (Lopes and Oden, 1987; Griffiths and Tenenbaum, 2003; Planton et al., 2021, 2022) and Koffka (2013)). A large body of research on similar sequences supports the generality of our compression operations.

K-complexity for license plates.

The k-complexity of a license plate is defined by the final number of digits after all possible operations of our language have been applied.[18]

We illustrate how an attribute can reduce the complexity of a number. Consider number 646. This is a palindrome and can be simplified using operation B. Once this simplification is made, the number can be fully characterized using only the first two digits: 64. Thus, the k-complexity of 646 equals that of 64, i.e., 2.[19]

For a given plate length, the k-complexity of a number is determined as follows. For numbers that cannot be simplified, their k-complexity corresponds to their number of digits. For three- and four-digit palindromes, their k-complexity is 2; for five-digit palindromes, it is 3. Similarly, pairs and doubles (e.g., 3344) have a k-complexity of 2 and necessarily consist of four-digit plates. Unique repeated numbers have a k-complexity of 1, while multiples of 100 and 1000 have k-complexities equal to their number of digits minus two and three, respectively.

Note that unique repeated numbers may also qualify as palindromes, pairs, or doubles. However, we focus on the operation yielding the lowest k-complexity. Overall, we construct a measure of the k-complexity of a number for a given plate length. Table 4 displays the distribution of license plates by k-complexity and the number of digits.

[18]We abstract from potential cognitive costs in performing simplifications, although we account for these in the empirical analysis by considering different categories for each type of simplification. Aksentijevic and Gibson (2012) and Aksentijevic et al. (2020) provide measures of complexity that account for the cognitive costs of different operations in binary sequences. Thus, the k-complexity of a number is independent of the number of required operations of simplification. Note that for all plates in our dataset, simplifications can be completed in a single step.

[19]On principle, several operations of simplifications can be applied successively. For instance, for number 64646, a first operation of simplification using the fact that it is a palindrome allows to fully characterize the number using only the first three digits: 646. Thus, the k-complexity of 64646 after this operation is that of 646, i.e., 3. The number can be further simplified by applying the same operator (B) to 646, reducing the complexity to that of 64, i.e., 2. As mentioned earlier, multiple operations do not take place in our data set as all numbers can be simplified using at most one operation.

Table 4: Distribution of plates by digits and k-complexity

	k-complexity					
Number of digits	1	2	3	4	5	Total
3	5	19	170			194
4	2	15	0	256		273
5	0	3	72	0	6,117	6,192
Total	7	37	242	256	6,117	6,659

Most plate numbers cannot be simplified, and, as expected, it is easier to simplify three-digit plates than five-digit plates.

Redundant information.

Another important factor for the compression of information is the quantity of redundant information. We capture redundancy by considering for each plate length the number of unique digits. For instance, operator E of our language corresponds to numbers composed of one unique digit and for which the degree of redundancy is the highest.[20]

This concept also relates to cognitive costs involved in processing different pieces of information, as explored in the literature on mental information processing (Miller, 1956).[21] Compression programs eliminate redundancy while using theoretical components to compress the informational content of a signal. In our analysis, we distinguish between the k-complexity, derived from our language, and the redundancy of a number.

Thus, some numbers have the same k-complexity but different redundancy levels. For instance, 23224 and 94783 have the same k-complexity, but 23224 has greater redundancy than 94783.

We construct variables to capture a number's redundancy. For all plates of a given length (three, four, or five digits), we create an interaction variable l-unique for numbers composed of l unique digits. For example, 911 is a three-digit plate with two unique digits, so it belongs to 2-unique. Similarly, 23224 is a five-digit number with three unique digits: 2, 3, and 4, and belongs to 3-unique. We assume redundancy is relevant only for numbers that cannot be compressed and thereby exhibit max k-complexity (i.e., plates with k-complexity equal to the number of digits). Table 5 shows the distribution of plates with max k-complexity by the number of unique digits.

[20] Enke (2020) uses a similar notion using redundancy to characterize the complexity of information spaces.

[21] This cognitive cost reflects the effects of mental representation described by the mental number line literature (Brysbaert, 1995). For example, Izard and Dehaene (2008) and Sobkow et al. (2019) identify cognitive costs in experiments where participants face novel numerical values.

Table 5: Distribution of max k-complexity plates by number of unique digits

	Number of unique digits				
Number of digits	2	3	4	5	Total
3	38	132			170
4	6	94	156		256
5	36	679	3,219	2,183	6,117
Total	80	905	3,375	2,183	6,543

This table includes only license plates with max k-complexity (i.e., plates with k-complexity equal to their number of digits). Numbers with only one unique digit are excluded since they have a k-complexity equal to one.

Simplicity vs. Scarcity.

It is important to stress that simplicity, as we define it, is distinct from probability and scarcity. A simple number or set of numbers is not necessarily rare. For instance, the proportion of two-digit palindromes between 0 and 99 is $\frac{9}{100}$. In our constructed language, palindromes are simpler than multiples of numbers like 37, even though the probability of randomly selecting a palindrome is higher than the probability of selecting a multiple of 37.

We can further illustrate this distinction by comparing palindromes of odd and even lengths. For instance, there are 90 palindromes with three and four digits. Although they share the same level of scarcity, our empirical analysis reveals that three-digit palindromes tend to have higher market values than four-digit palindromes. This is because shorter numbers typically have lower k-complexity. Put differently, the mental operation required to simplify four-digit palindromes is more cognitively demanding than for three-digit palindromes, which in turn results in lower auction prices for the former.

5 Analogical Complexity

Some signals may be theoretically complex but still easy for a receiver to process due to social and reputational factors. For instance, numbers associated with significant cultural or historical references can be quickly understood. In the case of license plates, numbers like historical dates or culturally significant references, such as 911 (the emergency number in the US), are frequently sold at higher prices because they are easily recognizable and meaningful to a wide audience.

These effects are well-documented in cognitive science where researchers have developed the concepts of analogical reasoning and relational complexity (Goswami, 1989; Halford, 1992). We borrow these concepts to measure how easily a number is understood or appreciated based on its societal significance. This analogical simplicity of a number, based on its social appreciation, is quantified using the Google Frequency Index, building on the work of Cilibrasi and Vitanyi (2007).

The Google Frequency Index.

To create the Google Frequency Index for license plates, we developed an algorithm that scrapes the number of search results returned by Google for each license plate number.[22]

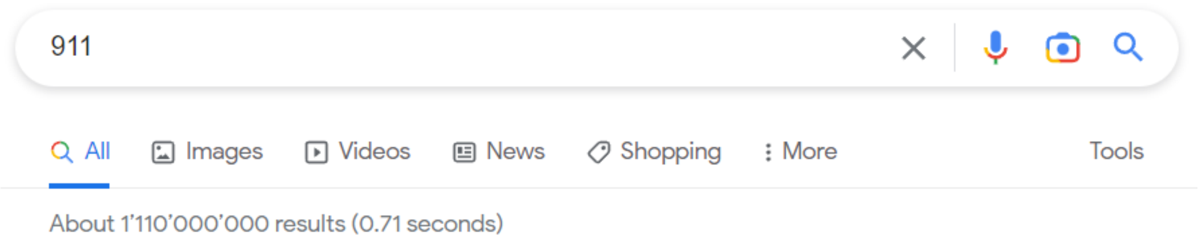

Figure 3: Google's results page for the query "911"

For instance, Figure 3 shows that the query "911" yields over 1 billion 110 million results, demonstrating a high level of social recognition. Conversely, numbers with fewer results are understood by fewer people and may carry less cultural weight.[23] The resulting table constitutes the index of analogical simplicity of the numbers in our dataset.[24]

We compute the Google Frequency Index for all license plate numbers in our dataset and use the frequency of search results as a proxy for analogical simplicity. The mean number of Google results across the dataset is 23.663 million, with a log-transformed mean of 14.167. Standard deviations are 387.763 million and 1.390, respectively, showing substantial variation in how culturally significant different numbers are.

This measure of analogical simplicity provides a valuable complement to the theoretical simplicity discussed earlier. Observers will understand more easily a signal if its analogical complexity is low enough, allowing them to refer to the informational content of the signal to its set of cultural references. A signal that ranks relatively low in the Google Frequency Index will be understood by a small number of people.

The methodology that we use to construct the Google Frequency Index can be applied in other contexts. For example, it could be used to rank the reputation of universities by querying the number of search results associated with each institution. This would allow for the creation of a reputation index based on public recognition, potentially adjusted for geographic or cultural variations by filtering search results based on the location of the query.

[22] Accessed and scraped on February 15, 2023.

[23] Note that analogical simplicity is contextual and depends on the culture shared by a group of people, compared to theoretical simplicity, which is universal.

[24] We also use for robustness the ranking of numbers according to their (inverse) log of queries, following Zipf's law (Zipf, 1942). The resulting variable has a much lower explanatory power than the frequency index.

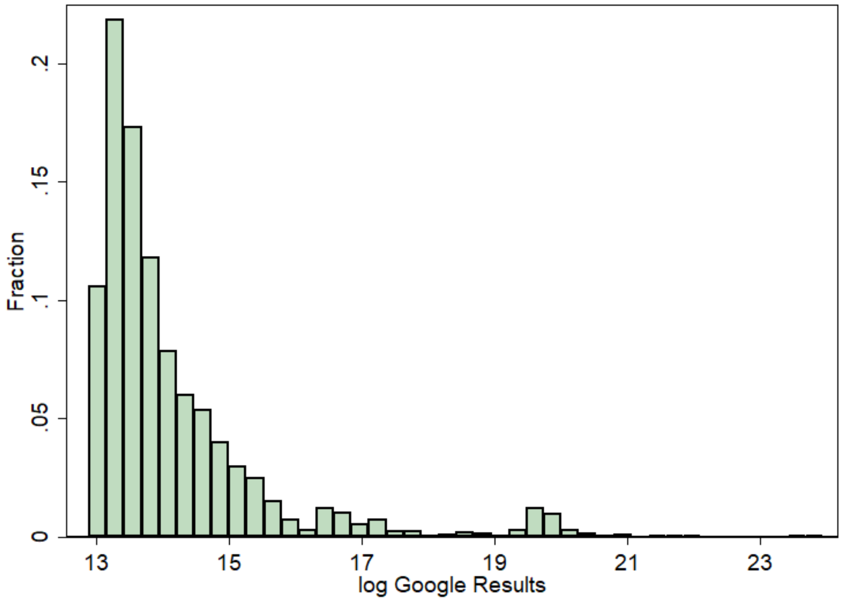

Figure 4: Histogram of log of Google search results.

6 A Model of Complex Signaling in Auctions

We provide intuitions of the main mechanisms at play by analyzing a model of auctions where the purchased good has a conspicuous value, and the resulting signal can be more or less complex. In turn, we show how the complexity of the signal impacts the willingness of auction participants to bid for the good.

Our model brings together elements from three different fields: the literature on conspicuous consumption, models of signaling in auctions, and models of decision-making with complex information.

6.1 Auctions

We consider an auction with n participants bidding for a single conspicuous good. Bidders differ in their wealth, denoted by θ, which is private information drawn from a distribution $F(.)$ on $[a, b]$. Bidders choose to allocate part of their wealth towards bidding in the auction, in line with classical models of conspicuous spending (e.g., Ireland (1994)). The utility from consuming a numeraire good, denoted by u, is assumed to be positive, increasing, and concave. Hence, a losing bidder's utility is $u(\theta)$.

The conspicuous good itself holds no intrinsic value but provides signaling value, consistent with the idea of "money burning" in conspicuous consumption (Bagwell and Bernheim, 1996). After the auction, outsiders observe who has acquired the good, which carries a signal simplicity denoted by γ, where higher γ corresponds to a simpler signal. The observer can infer the wealth type of the bidder based on this simplicity. The distribution $F(\cdot)$ and the number of bidders are common knowledge.

An observer expects the status S of the winning bidder to be the expected value of the largest of n draws from $F(\cdot)$:

$$S = E(\theta \mid \text{winner}) = \int_a^b x dF^n(x)$$

Thus, a bidder with wealth θ, bidding as if they were of type $\tilde{\theta}$, has the following expected utility:

$$\underbrace{\int_a^{\tilde{\theta}} u(\theta - \beta(x))dG(x)}_{\text{Base consumption (winning bid)}} + \underbrace{\int_{\tilde{\theta}}^b u(\theta)dG(x)}_{\text{Base consumption (losing bid)}} + \underbrace{\gamma \int_a^{\tilde{\theta}} SdG(x)}_{\text{Signaling value (winning bid)}}$$

The first-order conditions at $\tilde{\theta} = \theta$ characterize the equilibrium:

$$u(\theta - \beta(\theta))G'(\theta) - u(\theta)G'(\theta) + \gamma G'(\theta)S = 0,$$

leading to the equilibrium bid function:

$$\beta(\theta) = \theta - u^{-1}(u(\theta) - \gamma S)$$

A sufficient condition for the bid function to be positive is $u(a) > \gamma b$, ensuring $u(\theta) > \gamma S$.

Performing comparative statics on γ shows that *the equilibrium bid function increases with the simplicity of the product.* Additionally, we can show that the bid function increases with θ, confirming that *the wealthiest bidder wins the auction.*

Proposition 1 *Auction participants place bids that increase with simplicity γ. The wealthiest bidder wins the auction.*

6.2 Observer Utility

We now provide micro-foundations for γ, explaining how it relates to the ability of observers to process the signal depending on its complexity $\kappa > 1$. Observers derive utility 1 from making an inference, at a cognitive cost $c(\kappa) = c \cdot \kappa$, which increases with complexity. Let the agents' cognitive cost parameter c be uniformly distributed over $[0, 1]$.

Thus, the utility for an agent of type c from making an inference is:

$$v(c, \gamma) = 1 - c \cdot \kappa$$

In equilibrium, only a fraction $\frac{1}{\kappa}$ of the population interprets the signal. This yields a functional relationship between γ and complexity κ: $\gamma = \frac{1}{\kappa}$.

7 Empirical Analysis

7.1 Description of the Variables

In this section, we describe the variables used in the empirical analysis. We start with the two measures of complexity.

Analogical complexity. Analogical complexity is captured by the variable *Google-FI* representing the Google Frequency Index defined in section 5.

k-complexity. The dummy variable *i-k-complexity* ($i = 1, \ldots, 5$) equals 1 if the plate number has a k-complexity of i. For example, a four-digit plate number can have a k-complexity of 1 (for a unique digit), 2 (for palindromes or multiples of 100), or 4 (for numbers that cannot be simplified). Plate numbers with the highest k-complexity ($i = 5$) represent the baseline category in regression models.

Plate length. The dummy variable *j-digits* ($j = 3, \ldots, 5$) is equal to one if the plate number has j digits.

Number of unique digits. The dummy variable *l-uniques* ($l = 2, \ldots, 4$) equals one if the plate number has l unique digits.

Interactions. To capture the interaction between k-complexity and plate length, we create interaction variables *i-k-complexity#j-digits*, *j-digits#l-uniques*, and *j-digits#max-k-complexity#l-uniques*. These allow us to study how different complexity measures affect the value of the license plate. For instance, the k-complexity could have a different impact on price depending on the number of different digits in the plate. This type of variation is captured by our interaction variables.

7.2 Empirical Results

Using our auction data, we estimate the effect of simplicity on the price of license plates. We employ ordinary least squares (OLS) regression to analyze the relationship between the price of a plate and our two different measures of complexity – theoretical complexity and analogical complexity – as well as the degree of redundancy.

The effect of simplicity on the log price of a license plate is summarized in Table 6.[25] In column (1) we regress log prices on analogical complexity only, and in column (2) on theoretical complexity only. In column (3) we then regress log prices on both analogical and theoretical complexity. In column (4) we add digit redundancy, on top of the two other measures of complexity.

[25] We have also used the price in CHF as a dependent variable, but the R^2 is lower. The results are available upon request.

In columns (2) and (3) the baseline category includes 5-digit plates with a k-complexity of 5. *3-digit* and *4-digit* are intercept shifters, and, therefore, each variable captures the effect for three and four-digit plates that cannot be simplified, and for which the k-complexity is the highest (equal respectively to 3 and 4). Similarly, in column (4) the baseline category includes 5-digit plate numbers with a k-complexity of 5 and that have five different digits.

Table 6: Regression of log of Prices on Simplicity (all interactions)

	(1)		(2)		(3)		(4)	
log Price	coef	tstat	coef	tstat	coef	tstat	coef	tstat
log of Google-FI	0.448	(132.88)			0.277	(45.90)	0.275	(50.59)
3-digits			2.798	(89.86)	1.149	(25.53)	1.224	(28.54)
4-digits			1.923	(75.28)	1.006	(33.60)	1.051	(33.15)
3-digits#1-k-complexity			1.158	(6.37)	0.564	(3.55)	0.284	(1.55)
3-digits#2-k-complexity			0.287	(2.96)	0.302	(3.58)	0.349	(4.55)
4-digits#1-k-complexity			1.314	(4.62)	0.835	(3.37)	0.909	(4.08)
4-digits#2-k-complexity			0.595	(5.59)	0.442	(4.77)	-0.196	(0.88)
5-digits#2-k-complexity			1.436	(6.21)	0.477	(2.35)	0.594	(3.27)
5-digits#3-k-complexity			0.715	(15.07)	0.444	(10.62)	0.087	(1.31)
3-digits#1-k-complexity#1-uniques							0.829	(2.90)
3-digits#3-k-complexity#2-uniques							0.211	(3.66)
4-digits#2-k-complexity#2-uniques							0.819	(3.45)
4-digits#4-k-complexity#2-uniques							0.400	(3.07)
4-digits#4-k-complexity#3-uniques							0.169	(4.12)
5-digits#3-k-complexity#2-uniques							1.531	(12.43)
5-digits#3-k-complexity#3-uniques							0.500	(6.08)
5-digits#5-k-complexity#2-uniques							1.104	(20.98)
5-digits#5-k-complexity#3-uniques							0.447	(32.48)
5-digits#5-k-complexity#4-uniques							0.103	(11.92)
Constant	1.225	(25.49)	7.407	(1,446.45)	3.576	(42.80)	3.492	(46.29)
Observations	6,659		6,659		6,659		6,659	
R-squared	0.726		0.701		0.773		0.818	

t-statistics in parentheses.

Our results on k-complexity show that, using this notion, simpler plates reach a higher price than more complex ones. Reducing the number of digits and a lower k-complexity increases the price of a plate.

It is remarkable that the positive relationship between simplicity and price holds for all measures of complexity. Consider column (4). 3-digit plates carry a premium of 122% over

5-digit plates, while this premium is 105% for 4-digit plates. Now consider the first set of interaction variables (plate-length and k-complexity). For all plate lengths, plates with lower k-complexity have a higher price than the baseline category that corresponds to more complex plate numbers. In addition, the gradient is positive for any given plate length: 1-k-complexity plates are selling at a higher price than 2-k-complexity plates, which themselves sell at a higher price than 3-k-complexity plates.

We interpret the effect of the second set of interaction variables (plate length, k-complexity, and number of unique digits) on prices. For a given number of digits and k-complexity, the lower the number of unique digits, the higher the price of a plate. This effect is persistent across all pairs of *j-digits* and *i-k-complexity*. Moreover, we find a positive price gradient of simplicity measured here by the number of unique digits. For 5-digit plates with a k-complexity of 5, reducing the number of unique digits increases the value of the plate by 10% for 4 unique digits, 45% for 3 unique digits, and 110% for 2 unique digits compared to the baseline category. We observe similar effects for the other interaction variables, making this result extremely robust.

Turning to analogical complexity, we see that the elasticity of the price of a plate with respect to the Google Frequency Index is 0.275, meaning that the price of a plate increases by 27.5% when the value of the Google Frequency Index increases doubles. This effect is highly significant and highlights the importance of analogical simplicity to explain the value of a plate.

Running the regression including only the analogical complexity, or a set of variables characterizing the k-complexity (without analogical complexity), we see that the goodness of fit is similar in both cases (with $R^2 = 0.726$ in column 1 vs. $R^2 = 0.722$ in column 3, see Table 8 in the Appendix). Including all complexity variables increases the goodness of fit to $R^2 = 0.818$. Note also that the coefficient associated with analogical complexity decreases from 0.448 to 0.275.

To assess the contribution of each measure of complexity, we ran three f-tests. First, we test whether the Google Frequency Index is significant and we find an F-statistics of 2559.85 with a p-value of zero. Secondly, we test whether all variables related to k-complexity are significantly different from zero, and find an F-statistics of 179.07 with a p-value of zero. Finally, we test whether all variables related to the number of unique digits are significant, and find an F-statistics of 114.32 with a p-value of zero. Thus, all three measures of complexity are important determinants of the variance of the log of the price of a plate.

To summarize, analogical complexity and k-complexity are important factors determining the price of a license plate, as well as the number of unique digits. In general, a plate that is analogically simple also has a low k-complexity, but this does not hold for all plates. Thus, despite a high level of correlation, both concepts are complementary and allow to better predict the price of a plate.

8 Implications

Our results have interesting implications for firms and for scholars working in the field of complexity. The two measures of complexity we introduce are flexible and can be applied using straightforward tools tailored to the nature of the signal considered.

First, conspicuous motives may generate a significant price premium in luxury industries. The notion of brand prominence has been famously conceptualized by Han et al. (2010) to explain the conspicuous value of a good and the purchasing behavior of consumers of luxury goods. Our measures of complexity provide a cognitive theoretical grounding for this role of prominence, and can be used to assess the degree of prominence of a brand. Below we discuss a case study where a company can use the measures of complexity introduced in this article.

Consider a brand willing to evaluate the degree of prominence of a product, either one of its own products, or of a competitor. The brand can simply use our index of analogical complexity. Indeed, the degree of prominence of a brand relates directly to how simple it is for an observer to infer the price of a good from seeing the object, a notion captured by our analogical complexity. The Google Frequency Index can therefore directly be used by a brand to optimize its (competitive) pricing decision, and the measure of analogical complexity can be used to optimize competitive interactions between firms. Using variations of the index with IP addresses, a company can also assess the different degrees of prominence of its good and of those of competitors controlling for detailed geographical locations. Moreover, using Google trends, market analysts can evaluate variations of prominence in the industry for several years and better understand the competitive trends resulting from the conspicuous value of different luxury products.[26] Beyond luxury brands, universities or charities could also account for this role of complexity to optimize their interactions with consumers, prospective students, and donors.

Secondly, our results suggest that luxury brands should develop strategies for their goods to be the simplest possible signals for someone's status. For instance, luxury fashion house Facetasm is famous in Japan, where its clothes easily sell for thousands of dollars, but much less in other countries where the brand has not yet achieved the necessary reputation to allow consumers to signal their wealth by purchasing Facetasm's products.[27] For this brand

[26] Additionally, we can construct indexes based on images using image search tools, such as Google lens. This approach could be valuable for comparing the social significance of logos or the specific appearances of products.

[27] Similar examples include luxury fashion house Akris in Switzerland.

to export outside of Japan, it must be understood in other countries as a signal of wealthy consumption. To establish such reputation, Facetasm could enter into partnerships with luxury brands that are well established in other countries. By doing so, it can use the social significance of other brands to establish in turn its own simple status signal.

Thirdly, our methodology is relevant beyond signaling theory in various contexts. For instance, papers have analyzed the complexity of legal regimes (Ash et al., 2024), tax systems (Schuck, 1992; Slemrod, 2005; Bennani and Neuenkirch, 2023), regulations (Colliard and Georg, 2023), contracts (Jakobsen, 2020), or terms of agreements (Cranor et al., 2006) using case-specific definitions. These articles have adopted two types of approaches to characterize complexity.

On the one hand, authors can focus on the (uncompressed) complexity of an information set. For instance, Slemrod (2005) quantifies the complexity of a tax system simply by counting the number of lines in the tax forms and the number of pages of accompanying instructions. Cranor et al. (2006) characterizes the complexity of privacy agreements using the number of choices a user is required to consider.

On the other hand, other articles measure the complexity of an information set focusing on its compressed informational content. In the context of legal regimes, Ash et al. (2024) analyzes the semantic content of a text, while Jakobsen (2020) use a similar approach for contracts. Although the measure of k-complexity introduced in this article – characterized by the compressed quantity of information – may not outperform previous measures constructed on case-specific basis, we argue that the k-complexity still offers a valuable approximation of the informational content across various studies, with minimal additional costs for researchers. An author willing to characterize the complexity of a tax system could use readily available compression algorithms such as gzip, bzip2, or zlib, which provide efficient approximations (and almost costless in terms of time and processing costs) of the k-complexity of large files. Similar tools are available for instance in Python libraries gzip, lzma, or zlib. Compression operations performed by these tools eliminate repetitive patterns, allowing to focus on the semantically meaningful information. Hence, we believe our analysis provides a theoretical basis for using such compression tool to evaluate the complexity of information sets in a wide range of contexts.

9 Conclusion

This article contributes to the understanding of signaling behaviors by providing measures of the complexity of signals and by showing that senders prefer simple signals. Our measures of theoretical and analogical complexity offer a new way for researchers to characterize the complexity of various situations.

This analysis opens a range of questions for future research, which can be partially answered using the methodology presented in this article.

First, how general are the notions of theoretical and analogical complexity introduced in this article? Answering this question requires exploring applications to other types of signals. In other applications, theoretical complexity can be characterized using the procedure of Ziv and Lempel (1977) for large information sets, and the programs of Soler-Toscano et al. (2014) for short sequences with an alphabet of size up to 9, and of Aksentijevic et al. (2020) for binary sequences accounting for the complexity of different compression operations. Alternative indexes can also be used to characterize analogical complexity. For instance, for long-term reputation effects, cultural meaningfulness, and historical settings, the Google Books website provides a relevant source of information to construct the index. More generally, Google Trends can also be used to characterize the evolution of relative social meaningfulness and analogical complexity over time.

Secondly, what is the role of complexity in games where senders want to send a signal to only a subset of agents? For instance, in the status signaling games famously analyzed by Bourdieu (1982) people can send signals that are understood only by members of similar social classes (for instance, using cultural references). The analysis of signaling with selection of receivers using signal complexity is a topic for further research.[28]

Future research could also include the cognitive costs of processing signals to better understand the effects of signal selection and their efficiency.

[28] Jin et al. (2022) provide an interesting analysis in this direction, by showing that an agent can increase the complexity of a message to modify the behavior of the receiver, in a model of endogenous obfuscation.

References

Abeler, J., D. Huffman, and C. Raymond (forthcoming). Incentive complexity, bounded rationality and effort provision. *American Economic Review*.

Abeler, J. and S. Jäger (2015). Complex tax incentives. *American Economic Journal: Economic Policy 7*(3), 1–28.

Aksentijevic, A. and K. Gibson (2012). Complexity equals change. *Cognitive systems research 15*, 1–16.

Aksentijevic, A., A. Mihailovic, and D. T. Mihailovic (2020). Time for change: implementation of aksentijevic-gibson complexity in psychology. *Symmetry 12*(6), 948.

Ali, H. H., S. Lecocq, and M. Visser (2008). The impact of gurus: Parker grades and en primeur wine prices. *The Economic Journal 118*(529), F158–F173.

Amalric, M., L. Wang, P. Pica, S. Figueira, M. Sigman, and S. Dehaene (2017). The language of geometry: Fast comprehension of geometrical primitives and rules in human adults and preschoolers. *PLoS computational biology 13*(1), e1005273.

Ash, E., M. Morelli, and M. Vannoni (2024). More laws, more growth? evidence from us states. *Journal of Political Economy*.

Azeredo da Silveira, R., Y. Sung, and M. Woodford (2024). Optimally imprecise memory and biased forecasts. *American Economic Review 114*(10), 3075–3118.

Bagwell, L. S. and B. D. Bernheim (1996). Veblen effects in a theory of conspicuous consumption. *The American economic review*, 349–373.

Banovetz, J. and R. Oprea (2023). Complexity and procedural choice. *American Economic Journal: Microeconomics 15*(2), 384–413.

Barron, A., J. Rissanen, and B. Yu (1998). The minimum description length principle in coding and modeling. *IEEE transactions on information theory 44*(6), 2743–2760.

Bennani, H. and M. Neuenkirch (2023). Too complex to digest? federal tax bills and their processing in us financial markets. *International Tax and Public Finance*, 1–25.

Bernheim, B. D. (1994). A theory of conformity. *Journal of political Economy 102*(5), 841–877.

Bernheim, B. D. and C. Sprenger (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-independent probability weighting. *Econometrica 88*(4), 1363–1409.

Bos, O. and T. Truyts (2021). Auctions with signaling concerns. *Journal of Economics & Management Strategy 30*(2), 420–448.

Bourdieu, P. (1982). La distinction. *Revue Philosophique de la France et de l'Etranger 172*(1).

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of experimental psychology: General 124*(4), 434.

Calude, C. S. (2002). *Information and randomness: an algorithmic perspective.* Springer Science & Business Media.

Carvalho, L. and D. Silverman (2024). Complexity and sophistication. *Journal of Political Economy Microeconomics 2*(1), 43–76.

Célérier, C. and B. Vallée (2017). Catering to investors through security design: Headline rate and complexity. *The Quarterly Journal of Economics 132*(3), 1469–1508.

Charness, G. and D. Levin (2005). When optimal choices feel wrong: A laboratory study of bayesian updating, complexity, and affect. *American Economic Review 95*(4), 1300–1309.

Charness, G. and D. Levin (2009). The origin of the winner's curse: a laboratory study. *American Economic Journal: Microeconomics 1*(1), 207–236.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review 103*(3), 566.

Chater, N. and P. Vitányi (2003). Simplicity: a unifying principle in cognitive science? *Trends in cognitive sciences 7*(1), 19–22.

Cilibrasi, R. L. and P. M. Vitanyi (2007). The google similarity distance. *IEEE Transactions on knowledge and data engineering 19*(3), 370–383.

Colliard, J.-E. and C.-P. Georg (2023). Measuring regulatory complexity. *HEC Paris Research Paper No. FIN-2020-1358*.

Cranor, L. F., P. Guduru, and M. Arjula (2006). User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction (TOCHI) 13*(2), 135–178.

Cremer, J., L. Garicano, and A. Prat (2007). Language and the theory of the firm. *The Quarterly Journal of Economics 122*(1), 373–407.

Dehaene, S., F. Al Roumi, Y. Lakretz, S. Planton, and M. Sablé-Meyer (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences*.

Dewatripont, M. and J. Tirole (2005). Modes of communication. *Journal of Political Economy 113*(6), 1217–1238.

Dimson, E. and C. Spaenjers (2011). Ex post: The investment performance of collectible stamps. *Journal of Financial Economics 100*(2), 443–458.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics 135*(3), 1363–1398.

Enke, B., T. Graeber, and R. Oprea (forthcoming). Complexity and time. *Journal of the European Economic Association*.

Enke, B. and F. Zimmermann (2019). Correlation neglect in belief formation. *The Review of Economic Studies 86*(1), 313–332.

Foarta, D. and M. Morelli (2022). Complexity and the reform process: The role of delegated policymaking. *BAFFI CAREFIN Centre Research Paper* (2022-180).

Fodor, J. A. (1975). *The language of thought*, Volume 5. Harvard university press.

Fudenberg, D. and I. Puri (2022). Simplicity and probability weighting in choice under risk. In *AEA Papers and Proceedings*, Volume 112, pp. 421–25.

Gabaix, X. and T. Graeber (2023). The complexity of economic decisions. *Available at SSRN*.

Gauvrit, N., H. Singmann, F. Soler-Toscano, and H. Zenil (2016). Algorithmic complexity for psychology: a user-friendly implementation of the coding theorem method. *Behavior research methods 48*, 314–329.

Gayer, G. and I. Gilboa (2014). Analogies and theories: The role of simplicity and the emergence of norms. *Games and Economic Behavior 83*, 267–283.

Gell-Mann, M. and S. Lloyd (1996). Information measures, effective complexity, and total information. *Complexity 2*(1), 44–52.

Ghent, A. C., W. N. Torous, and R. I. Valkanov (2019). Complexity in structured finance. *The Review of Economic Studies 86*(2), 694–722.

Giovannoni, F. and M. Makris (2014). Reputational bidding. *International Economic Review 55*(3), 693–710.

Glanzer, M. and W. H. Clark (1963). Accuracy of perceptual recall: An analysis of organization. *Journal of Verbal Learning and Verbal Behavior 1*(4), 289–299.

Goeree, J. K. (2003). Bidding for the future: Signaling in auctions with an aftermarket. *Journal of Economic Theory 108*(2), 345–364.

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive development 4*(3), 251–268.

Griffiths, T. L. and J. B. Tenenbaum (2003). Probability, algorithmic complexity, and subjective randomness. In *Proceedings of the annual meeting of the cognitive science society*, Volume 25.

Grunwald, P. and P. Vitányi (2004). Shannon information and kolmogorov complexity. *arXiv preprint cs/0410002*.

Haile, P. A. (2003). Auctions with private uncertainty and resale opportunities. *Journal of Economic theory 108*(1), 72–110.

Halford, G. S. (1992). Analogical reasoning and conceptual complexity in cognitive development. *Human Development 35*(4), 193–217.

Hammer, D., A. Romashchenko, A. Shen, and N. Vereshchagin (2000). Inequalities for shannon entropy and kolmogorov complexity. *Journal of Computer and System Sciences 60*(2), 442–464.

Han, Y. J., J. C. Nunes, and X. Drèze (2010). Signaling status with luxury goods: The role of brand prominence. *Journal of marketing 74*(4), 15–30.

Hansen, M. H. and B. Yu (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association 96*(454), 746–774.

Hartmann, D., M. R. Guevara, C. Jara-Figueroa, M. Aristarán, and C. A. Hidalgo (2017). Linking economic complexity, institutions, and income inequality. *World development 93*, 75–93.

Hertel, J. and J. Smith (2013). Not so cheap talk: Costly and discrete communication. *Theory and decision 75*, 267–291.

Hidalgo, C. A. (2009). The dynamics of economic complexity and the product space over a 42 year period. *CID Working Paper Series*.

Hidalgo, C. A. and R. Hausmann (2009). The building blocks of economic complexity. *Proceedings of the national academy of sciences 106*(26), 10570–10575.

Hörner, J. and N. Sahuguet (2007). Costly signalling in auctions. *The Review of Economic Studies 74*(1), 173–206.

Ireland, N. J. (1994). On limiting the market for status signals. *Journal of public Economics 53*(1), 91–110.

Iyengar, S. S. and E. Kamenica (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics 94*(7-8), 530–539.

Izard, V. and S. Dehaene (2008). Calibrating the mental number line. *Cognition 106*(3), 1221–1247.

Jäger, G., L. P. Metzger, and F. Riedel (2011). Voronoi languages: Equilibria in cheap-talk games with high-dimensional types and few signals. *Games and economic behavior 73*(2), 517–537.

Jakobsen, A. M. (2020). A model of complex contracts. *American Economic Review 110*(5), 1243–73.

Jin, G. Z., M. Luca, and D. Martin (2022). Complex disclosure. *Management Science 68*(5), 3236–3261.

Juola, P. (2008). Assessing linguistic complexity. *Language Complexity: Typology, Contact, Change. John Benjamins Press, Amsterdam, Netherlands*.

Kemp, C. (2012). Exploring the conceptual universe. *Psychological review 119*(4), 685.

Koffka, K. (2013). *Principles of Gestalt psychology*, Volume 44. Routledge.

Kolmogorov, A. N. (1963). On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A*, 369–376.

Kurtz-David, V., A. Brandenburger, and P. Glimcher (2024). The limits of social cognition: Production functions and reasoning in strategic interactions.

Lempel, A. and J. Ziv (1976). On the complexity of finite sequences. *IEEE Transactions on information theory 22*(1), 75–81.

Leung-Yan-Cheong, S. and T. Cover (1978). Some equivalences between shannon entropy and kolmogorov complexity. *IEEE Transactions on Information Theory 24*(3), 331–338.

Li, M., P. Vitányi, et al. (2008). *An introduction to Kolmogorov complexity and its applications*, Volume 3. Springer.

Lim, H. E., W. C. Wong, and N. A. H. Abdullah (2017). The economics of luckiness: how crazy are malaysian chinese towards car plate number "8". *Malaysian Management Journal 21*, 49–60.

Lloyd, S. (2001). Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine 21*(4), 7–8.

Lopes, L. L. and G. C. Oden (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition 13*(3), 392.

Maćkowiak, B., F. Matějka, and M. Wiederholt (2023). Rational inattention: A review. *Journal of Economic Literature 61*(1), 226–273.

Mador, G., D. Sonsino, and U. Benzion (2000). On complexity and lotteries' evaluation–three experimental observations. *Journal of Economic Psychology 21*(6), 625–637.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review 63*(2), 81.

Murena, P.-A., M. Al-Ghossein, J.-L. Dessalles, A. Cornuéjols, et al. (2020). Solving analogies on words based on minimal complexity transformation. In *IJCAI*, pp. 1848–1854.

Oprea, R. (2020). What makes a rule complex? *American economic review 110*(12), 3913–3951.

Oprea, R. (2024). Decisions under risk are decisions under complexity. *American Economic Review 114*(12), 3789–3811.

Piantadosi, S. T., J. B. Tenenbaum, and N. D. Goodman (2012). Bootstrapping in a language of thought: A formal model of numerical concept learning. *Cognition 123*(2), 199–217.

Planton, S., F. Al Roumi, L. Wang, and S. Dehaene (2022). Compression of binary sound sequences in human working memory. *bioRxiv*, 2022–10.

Planton, S., T. van Kerkoerle, L. Abbih, M. Maheu, F. Meyniel, M. Sigman, L. Wang, S. Figueira, S. Romano, and S. Dehaene (2021). A theory of memory for binary sequences:

Evidence for a mental compression algorithm in humans. *PLoS computational biology 17*(1), e1008598.

Puri, I. (2024). Simplicity and risk. *Journal of Finance*.

Riley, J. G. (2001). Silver signals: Twenty-five years of screening and signaling. *Journal of Economic literature 39*(2), 432–478.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica 14*(5), 465–471.

Salant, Y. and J. L. Spenkuch (2022). Complexity and choice. Technical report, National Bureau of Economic Research.

Schuck, P. H. (1992). Legal complexity: Some causes, consequences, and cures. *DuKE Lj 42*, 1.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal 27*(3), 379–423.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99–118.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics 50*(3), 665–690.

Slemrod, J. (2005). The etiology of tax complexity: Evidence from us state income tax systems. *Public Finance Review 33*(3), 279–299.

Sobel, J. (2012). Complexity versus conflict in communication. In *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6. IEEE.

Sobkow, A., K. Fulawka, P. Tomczak, P. Zjawiony, and J. Traczyk (2019). Does mental number line training work? the effects of cognitive training on real-life mathematics, numeracy, and decision making. *Journal of Experimental Psychology: Applied 25*(3), 372.

Soler-Toscano, F., H. Zenil, J.-P. Delahaye, and N. Gauvrit (2014). Calculating kolmogorov complexity from the output frequency distributions of small turing machines. *PloS one 9*(5), e96223.

Sonsino, D., U. Benzion, and G. Mador (2002). The complexity effects on choice with uncertainty–experimental evidence. *The Economic Journal 112*(482), 936–965.

Spence, M. (1978). Job market signaling. In *Uncertainty in economics*, pp. 281–306. Elsevier.

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American economic review 92*(3), 434–459.

Szmrecsanyi, B. (2016). An informationtheoretic approach to assess linguistic complexity. *Complexity, isolation, and variation 57*, 71.

Thakor, R. T. and R. C. Merton (2023). Trust, transparency, and complexity. *The Review of Financial Studies 36*(8), 3213–3256.

Veblen, T. (1899). *The theory of the leisure class.*

Waelbroeck, P. (2003). Innovations, production complexity and the optimality of r&d. *Economics Letters 79*(2), 277–282.

Woodford, M. (2001). Imperfect common knowledge and the effects of monetary policy.

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics 12*, 579–601.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive psychology 24*(2), 220–251.

Yuan, Y., M. E. Deily, and Y. Yao (2022). Willingness to pay for status signals in online luxury markets. *Production and Operations Management 31*(2), 668–680.

Zenil, H., N. A. Kiani, and J. Tegnér (2017). Low-algorithmic-complexity entropy-deceiving graphs. *Physical Review E 96*(1), 012308.

Zipf, G. K. (1942). The unity of nature, least-action, and natural social science. *Sociometry 5*(1), 48–62.

Ziv, J. and A. Lempel (1977). A universal algorithm for sequential data compression. *IEEE Transactions on information theory 23*(3), 337–343.

A Appendix

A.1 Evidence of Vanity Plates Around the World

We provide here references for the examples of Section 3.

- United Arab Emirates: Abu Dhabi License Plate Fetches $14 Million, Sets World Record, Bloomberg, February 16, 2008.
- The UK: Personalised number plates, and Rich People's Problems: Are personalised number plates naff? The Financial Times, February 22, 2018.
- Czech Republic: Personalized car license plates to be available soon, radio Prague international, December 29, 2015.
- Norway: 2000 applications for personalised number plates in 25 minutes, NRK, June 15, 2017.
- Canada: Vanity licence plates coming to Quebec, CBC News, January 30, 2014.
- China: Auctioning license plates far better than dodgy lottery, China.org, May 4, 2011.
- Taiwan: Taiwan's vanity plates sell for big bucks, Taipei times, April 26, 2022.
- Hong-Kong: Hong Kong Just Auctioned Off a License Plate for 2.3 Million USD, Bloomberg, February 23, 2015.
- Malaysia: Lim et al. (2017).
- The US: Vanity License Plates Are Now a Million Dollar Business, The Street, February 23, 2023.
- Milton Friedman: Remarks by John B. Taylor at the Milton Friedman Memorial, Econbrowser, last accessed March 1, 2023.

A.2 Data Description

License plates are sold through online second-price auctions (ascending bids) with publicly observable bids. If a bid is placed in the last five minutes of the auction, there is an automatic five-minute extension to the auction. Bidders put the highest amount they are willing to bid and the website automatically over-bids until its limit is reached (this feature is also present on eBay for instance). In case of a tie, the bid submitted the earliest wins.

The auction website provides a summary of the characteristics of a plate: number on the plate; and its corresponding attributes – palindrome, potential date, the number of digits, and only one unique digit.

A.2.1 The Configuration of Listing

Each auction includes a reserve price and a minimum bid increment. These are assigned solely based on the number's numerical value and are not affected by its attributes. Presented in Table 7, we see that three, four and six-digit plates are auctioned at reserve prices of 4000, 2000 and 100 CHF, respectively. Five-digit plates are auctioned starting at 500, 300 and 200 CHF with thresholds at the plate numbers 20000 and 50000, respectively. The minimal bid increments are each set at one-tenth of the respective reserve price.

Table 7: Assignment of reserve price and minimal bid increment

	Number of digits on plate					
	3	4	5 (10000-19999)	5 (20000-49999)	5 (50000-99999)	6
Reserve price	4000	2000	500	300	200	100
Minimal bid increments	400	200	50	30	20	10

A.2.2 Variables

Palindromes.

Variable *cpalindrome* was constructed in Excel using the reverse function that inverses the digits of a number. The variable then returns 1 if the reversed number is equal to the original number, and zero else.

Length of a plate.

Variable *clength_plate* is constructed in Excel and returns the number of digits of each plate.

Unique repeated number.

Variable *cunique_repeated_number* is constructed in Excel and returns 1 if a number is constituted of only one digit repeated several times. It can be obtained directly using variable *cunique_repeated_number* = 1.

A.3 Additional regression

Table 8: Regression of log of Prices on Simplicity

	(1)		(2)		(3)		(4)	
log Price	coef	tstat	coef	tstat	coef	tstat	coef	tstat
log of Google Res.	0.448	(132.88)					0.270	(49.32)
3-digits			2.798	(89.86)	2.865	(86.94)	1.252	(28.90)
4-digits			1.923	(75.28)	1.974	(64.77)	1.067	(33.28)
3-digits#1-k-complexity			1.158	(6.37)			0.625	(4.32)
3-digits#2-k-complexity			0.287	(2.96)			0.349	(4.49)
4-digits#1-k-complexity			1.314	(4.62)			0.917	(4.06)
4-digits#2-k-complexity			0.595	(5.59)			0.517	(6.03)
5-digits#2-k-complexity			1.436	(6.21)			0.610	(3.31)
5-digits#3-k-complexity			0.715	(15.07)			0.561	(14.62)
3-digits#max-k-complexity#2-uniques					0.210	(3.11)	0.211	(3.61)
4-digits#max-k-complexity#2-uniques					0.360	(2.35)	0.399	(3.03)
4-digits#max-k-complexity#3-uniques					0.150	(3.13)	0.168	(4.07)
5-digits#max-k-complexity#2-uniques					1.163	(18.82)	1.105	(20.74)
5-digits#max-k-complexity#3-uniques					0.465	(28.80)	0.447	(32.09)
5-digits#max-k-complexity#4-uniques					0.105	(10.32)	0.103	(11.78)
Constant	1.225	(25.49)	7.407	(1,446.45)	7.293	(926.74)	3.557	(46.76)
Observations	6,659		6,659		6,543		6,659	
R-squared	0.726		0.701		0.722		0.813	

t-statistics in parentheses. Column (3) only includes plates with max k-complexity: Baseline is 5-digit plate, max-k-complexity & 5 different digits.