CONSUMER FINANCIAL PROTECTION BUREAU | APRIL 2024

Price Complexity in Laboratory Markets

Prepared by Dustin Beckett

Consumer Financial Protection Bureau

Table of contents

Executive Summary

The paper reports on two laboratory experiments that investigated the effects of price complexity on market outcomes. The experiments had human participants interact in basic markets. Across experiment sessions, the experiments varied how complex sellers were allowed to make their prices. In treatment 1, sellers could describe their price only using a single number. In treatment 2, sellers described their products using up to eight "sub-prices," prices that added up to one total price. And, in treatment 3, sellers could use up to 16 sub-prices.

1. **Allowing more complex pricing led sellers to ask for higher total prices, on average.** Comparing one price and 16 sub-price markets, average seller asks increased by more than 60 percent.

2. **Allowing more complex pricing led buyers to make more mistakes, on average.** Comparing one price and 16 sub-price markets, the frequency of buyer mistakes increased by more than 16 percentage points, or almost 1,400 percent.

3. **Allowing more complex pricing led to higher transaction prices, on average.** Comparing one price and 16 sub-price markets, average prices increased by more than 70 percent.

4. **In markets with multiple sub-prices, sellers can compete for sales by increasing their price complexity to cause buyers to make mistakes.** In markets with multiple sub-prices, standard equilibria apparently do not hold. Models fit using experiment data show that when one seller chooses a low price, the other seller, instead of choosing an even lower price, can profit most by choosing a higher price with high complexity. In contrast, in the one price treatment, sellers cannot increase complexity and therefore most of the standard equilibria hold.

5. **In a second experiment, increasing market competition generally improved, but did not eliminate, the negative effects of complexity.** In markets with two sellers, average seller asks were about 64 percent higher in markets with more complexity. In markets with four sellers, average seller asks were only about 36 percent higher in more complex markets. In markets with two sellers, transaction prices were about 69 percent higher in markets with more complexity. In markets with four sellers, transaction prices were only about 38 percent higher in more complex markets. However, buyers made *more* mistakes in markets with more sellers. In markets with two sellers, buyers made mistakes about 9.9 percent of the time. In markets with four sellers, buyers made mistakes about 20.1 percent of the time.

1. Introduction

In the realm of consumer finance, consumers rarely face simple options. From frequent choices like how to pay for groceries, to the infrequent, like how to finance a home, consumers' choices are multifaceted. How well consumers navigate these choices can have large financial consequences, and many financial choices have therefore been well-studied. This type of scholarship generally focuses on tangible costs and benefits—fees, interest rates, rewards, etc.—and how well consumers navigate their choices given these costs and benefits. However, the *complexity of financial decision-making itself*, and the effects of complexity on consumers, are less often the direct object of study.

In this paper, the Consumer Financial Protection Bureau's (CFPB) Office of Research considers the effects of price complexity in a competitive market. We focus on a type of price complexity we call "dimensional complexity." Dimensional complexity is a measure of the complexity consumers face when they consider the price of a single product. A product's price has more dimensional complexity when its price is partitioned into multiple parts (e.g., as a mortgage has both upfront and interest costs) or if its use results in mandatory fees (e.g., as payment products such as checking accounts or prepaid cards may charge ATM or overdraft fees).[1]

It can be difficult to measure how price complexity affects consumers using naturally occurring data. One reason for this is that price complexity often covaries with product complexity. For example, if a product has additional fees, then it typically also has additional features. This means that we cannot simply compare products with different levels of price complexity and hope to conclude that differences between consumer outcomes are due to price complexity alone—differences could also arise from the differences in products.[2]

Another difficulty is that consumers typically buy products in the context of a broader marketplace. In markets, consumers and firms do not act in isolation—one consumer's actions affect others' outcomes, and their own outcomes are affected by the actions of other consumers and firms. For example, when a market has few sellers, then prices might be high, regardless of

[1] Of course, dimensional price complexity is but one type of complexity relevant to consumer choice. In section 2, we consider another type of price complexity we call "comparative complexity," a market-level measure of the complexity consumers face when they compare prices across sellers. Complexity could also arise from other sources not considered in this report, including complexity of products themselves (e.g., products with many features or products with features that are difficult to comprehend, use, or compare). Other examples from the literature include lack of evaluability (Hsee and Zhang, 2010), price complexity from difficulty in evaluating different parts of a product's price (Kalayci, 2016), or complexity from the need to trade-off between current and future costs or benefits (Chernev et. al, 2015).

[2] Moreover, consumers do not randomly choose products and therefore such an analysis could reflect selection bias.

the price complexity of the product. Conversely, if there are many sellers in a market, competition between sellers may result in low prices, even if the products' prices are complex.

In this paper, we report on two laboratory experiments that measure the effects of price complexity on consumers while attempting to account for the above difficulties. The laboratory allowed us to experimentally manipulate the dimensional complexity of products in simple markets in a way that would not have been possible in the real world.[3] In addition, the laboratory setting allowed us to control features of the market that would have been otherwise hidden—for example, consumers in our studies received the same benefits from purchasing a product, regardless of the product's price complexity. Finally, the lab allowed us to manipulate the number of sellers participating in markets to test how competition alters the effects of complexity.

For these experiments, CFPB staff collected data at the Gettysburg Lab for Experimental Economics (GLEE) in three distinct experiments between 2015 and 2019.[4] Both Experiment 1 and Experiment 2 experimentally manipulated the maximum dimensional complexity of products in the marketplace.[5] These experiments were identical except Experiment 2 clarified instructions and made improvements to participants' user-interfaces. The results from these experiments are similar, and so, for concision, we do not report the results of Experiment 1. For ease of description, we call Experiment 2 the "Complexity Experiment." In the next subsection, we describe related literature. Section 2 of this report describes this experiment. Subsections describe the theoretical model, experimental methodology, outcome measures and hypotheses, data, and experimental findings. Section 3 describes Experiment 3 which, in addition to manipulating price complexity, also experimentally manipulated the number of sellers operating in each market. For ease of description, we call this the "Competition Experiment." Section 4 offers concluding thoughts.

[3] Our experimental markets were modified versions of Bertrand Duopolies, two-seller markets in which sellers sell products that are identical except for their price, and sellers therefore compete for buyers through price alone. See section 2.

[4] Gettysburg College was awarded two separate contracts to complete this work under contract vehicle CFP-15D-00004 (Task Orders 0001 and 0003). Contract award amounts included participant payments. See, https://www.usaspending.gov/award/CONT_AWD_CFP15D000040001_955F_CFP15D00004_955F and https://www.usaspending.gov/award/CONT_AWD_CFP15D000040003_955F_CFP15D00004_955F

[5] As described in detail in section 2, participants acting as sellers were able to choose the actual level of dimensional complexity of their prices. Nevertheless, sellers generally used the complexity available to them, as realized dimensional complexity was highly correlated with the experimentally manipulated maximum dimensional complexity.

1.1 Related Literature

Carlin (2009) describes a model similar to the markets we explore in this paper. Although Carlin does not consider dimensional complexity directly, he conceives of a general price complexity and assumes that greater price complexity leads to a greater proportion of uninformed buyers. Implications of his model include: that greater price complexity leads to a greater frequency of buyer mistakes; that sellers who choose higher prices also choose greater price complexity; and that, in equilibrium, sellers will neither set prices equal to their marginal cost nor will they all set the same price (as in standard equilibrium characterizations of similar models).

The paper closest to this report is Kalaycı (2015), who sets out to test certain assumptions and implications of Carlin's model.[6] Like Carlin, Kalaycı does not consider dimensional complexity explicitly, and so his notion of complexity mixes dimensional complexity with mathematical difficulty. Other important differences are that Kalaycı allows for dimensional complexity to vary only between one and three sub-prices. In this paper, we allow dimensional complexity to vary between one and 16 sub-prices, more closely matching the range of dimensional complexity found in consumer financial products in the field (although in some cases still falling well short!). Another important distinction is that we experimentally vary the maximum dimensional complexity in its markets, enabling causal inference about the effects of price complexity. Kalaycı finds that buyers are more likely to make mistakes when sellers choose complex prices, and he finds limited evidence that sellers who set higher prices are more likely to choose greater price complexity.

This paper is related to several other literatures. Much has been written about "choice overload," an umbrella term used to describe negative effects on consumer choice of consumers having many options. Recent meta-analyses suggest that the mean effect of choice overload could be zero (Scheibehenne et al., 2010), or that the existence of an effect depends highly on the context in which choice overload is tested (Chernev et al., 2015). Nevertheless, a more recent paper (Dean et al., 2022), states that the apparent inconclusiveness in the choice architecture literature is due to the fact that most tests of choice overload are under-powered, and suggests that choice overload may be much more prevalent than previously thought.

In addition, our conception of dimensional complexity is similar to "partitioned pricing," the practice of dividing a product's total price into two or more mandatory sub-prices. A literature review of partitioned pricing literature shows the effects of partitioned pricing are varied and highly dependent on contextual and implementation details (Voester et al., 2016). Partitioned pricing can lead consumers towards more positive perceptions of products, greater purchase

[6] See also Kalaycı and Potters, 2011, and Kalaycı, 2016.

intentions, or greater willingness to pay, but these effects can be mitigated or even reversed depending on the details of implementation.[7] Nevertheless, in a recent natural field experiment, Carpenter et al. (2021) show that consumers are more likely to choose dominated financial products when those products' price disclosures are partitioned rather than consolidated.[8] White et al. (2023) replicate this result across a variety of product types.

Several papers have demonstrated that, strategically, whether it is a profitable strategy for firms to obfuscate prices (as via complexity) depends on how consumers react to obfuscation. If consumers infer that only firms with something to hide would obfuscate, then consumers would not buy obfuscated products and obfuscation would not be a profitable strategy (e.g., Viscusi 1978, Grossman 1981, Milgrom 1981). On the other hand, others have shown that if, for some reason, consumers do to make this inference, then obfuscation can be profitable (e.g., Carlin 2009, Gabaix and Laibson 2006, Brown et al. 2012). Indeed, Jin et al. (2019) recently showed that even in a very simple controlled environment, consumers make systematic mistakes when assessing complex disclosures and therefore obfuscation can be a profitable strategy for firms.

---

[7] For a discussion of partitioned pricing literature in the context of hotel resort fees, see https://www.ftc.gov/system/files/documents/reports/economic-analysis-hotel-resort-fees/p115503_hotel_resort_fees_economic_issues_paper.pdf

[8] A product is "dominated" if it is the same or worse than another product in every way and strictly worse in at least one way.

2. Complexity Experiment

2.1 The Model

Participants in the Complexity Experiment interacted in simple markets in which two participants acted as sellers and competed to sell homogeneous products (that is, there were no differences between the sellers' products) to two participants acting as buyers. Since the products were homogeneous, sellers could only differentiate their products via their prices. These markets therefore exhibited Bertrand competition (Bertrand, 1883). Since there are two sellers in these markets, they are also examples of Bertrand Duopoly.

In classic economic modelling of Bertrand competition, each seller, or firm, sets its price to maximize profits. Since products offered by the two firms are homogeneous, buyers view both products as perfect substitutes, and therefore buy from the lower-priced seller. Sellers, competing for sales, set their prices equal to their marginal cost of production and make no profits.

Sellers choosing prices equal to their marginal cost of production is a "Nash Equilibrium" of the Bertrand Duopoly model. A Nash Equilibrium is a set of strategies such that when all actors choose those strategies (and know others are playing them too), no single actor has an incentive to change their own strategy. Nash Equilibria are convenient for study because, in theory, behavior at Nash Equilibria is stable and predictable.

When prices are continuous, as they often are in theoretical considerations of Bertrand Duopoly, there is only one symmetric Nash Equilibrium of a Bertrand Duopoly.[9] When prices are discrete, as they are in these experiments, then there are two additional symmetric Nash Equilibria. Regardless, these other equilibria are very similar to the Nash Equilibrium described above as, in each of these equilibria, sellers choose prices just above their marginal cost of production.[10]

---

[9] A *symmetric* Nash Equilibrium is a Nash Equilibrium in which actors who have the same role (i.e., they have the same incentives, same set of possible actions, etc.) chose the same strategies.

[10] If we assume that when faced with two sellers with identical prices that buyers choose each seller with some positive probability, then, in addition to the equilibrium described above, it is a Nash Equilibrium for both sellers to set prices exactly one unit (e.g., one cent) higher than their marginal cost of production. In this equilibrium, sellers make small profits in expectation, and cannot undercut the other seller without losing these expected profits. If we further assume that when faced with two sellers with identical prices that buyers choose each seller with exactly 50% probability, then it is also an equilibrium for both sellers to set prices exactly two units (e.g., two cents) higher than their marginal cost of production. In this equilibrium, sellers make small profits in expectation, and cannot increase their expected payoffs by undercutting the other seller.

We chose to study markets with Bertrand competition for several reasons. First, these markets have clear equilibrium predictions. Second, these markets use simple, homogeneous goods, and thus allow us to control *product* complexity. And finally, these markets are easily modified to include additional price complexity in a way that does not substantively alter their Nash Equilibria.

2.2 The Experiment

In this experiment, we administered three different versions of Bertrand Duopoly. The first version was the standard Bertrand Duopoly. In this version, participants were randomly assigned to markets for homogeneous "objects" in which two participants were randomly assigned to be sellers and two participants were randomly assigned to be buyers. "Tokens" were the medium of exchange and sellers could set prices anywhere from 0 to 800 tokens, in whole token denominations. To represent the utility of consumption, and to keep product complexity constant, buyers received a fixed amount of 800 tokens for purchasing any object.[11] Sellers' cost of producing an object (their "marginal cost of production") was 100 tokens, and sellers only produced what they sold (that is, there was no risk of under- or over-production). As described in the previous section, there are three symmetric Nash Equilibria of this game, all very similar to one another, and each entail that both sellers set their prices at, or just above, their marginal cost of production of 100 tokens.[12] We call this version the "one price treatment."

The second version of the markets was identical to the first, but with one important difference. Rather than compete by setting a single price, sellers competed by setting eight "sub-prices."[13] Sub-prices were prices that added up to the total price that buyers would pay—there was nothing significant about the sub-prices themselves.[14] Buyers saw only sub-prices—the total price was not presented to them—but they could calculate the total price by summing the sub-prices.

[11] This also guaranteed that buyers had no chance of experiencing a loss in any round.

[12] The three symmetric Nash Equilibria are (1) both sellers set their prices at 100 tokens, (2) both sellers set their prices at 101 tokens, or (3) both sellers set their prices at 102 tokens.

[13] This is similar to the concept of "partitioned prices" found in other work. See, for example, Morwitz et al. (1998).

[14] Sellers could choose to either manually enter each of the eight sub-prices themselves, or they could start with a total price and press a button labeled "distribute" that randomly distributed the total price tokens amongst the sub-prices. In either case, the sum of the sub-prices, the total price, was displayed to sellers and updated whenever they changed a sub-price. The distribute button is one of the innovations made after Experiment 1. In Experiment 1, we found that sellers spent a long time entering sub-prices, and so, in Experiment 2, we offered the distribute button to help sellers accomplish their task more quickly. The behavior of the distribute button was modeled from sub-price setting behavior from Experiment 1. That is, for a given total price, we attempted to match the variance amongst the sub-prices that we observed sellers using in Experiment 1. Use of the distribute button was not limited, and sellers could achieve new distributions of sub-prices by pressing the distribute button subsequent times. Moreover, if a seller used the distribute button, she could still edit the sub-price values manually.

Sellers were aware that buyers would see only sub-prices. Because this version of the game is substantively equivalent to the first, the Nash Equilibria are also substantively equivalent. The Nash Equilibria of this version of the game involves sellers choosing the same *total prices* as in the Nash Equilibria of the first version. We call this version the "eight sub-price treatment."

Finally, the third version of the markets was identical to the second, but with 16 sub-prices instead of eight. The Nash Equilibria of this version of the game are also substantively equivalent in that sellers choose the same *total prices* as in the Nash Equilibria of the other two versions of the game. We call this version the "16 sub-price treatment."

FIGURE 1

Examples of buyers' choice interfaces across treatments

	SELLER 1	SELLER 2
Price	325	375
Choose:	○ Seller 1	○ Seller 2

SUB-PRICE	SELLER 1	SELLER 2
1	57	46
2	65	91
3	3	20
4	1	64
5	40	48
6	40	45
7	37	32
8	82	29
Choose:	○ Seller 1	○ Seller 2

SUB-PRICE	SELLER 1	SELLER 2
1	26	7
2	4	22
3	2	0
4	39	35
5	55	16
6	44	0
7	34	31
8	1	41
9	9	28
10	10	4
11	29	2
12	0	81
13	26	32
14	0	0
15	23	17
16	23	59
Choose:	○ Seller 1	○ Seller 2

The first panel displays an example of a buyer's choice in a market with one price. The second and third panels show examples in markets with eight and 16 sub-prices, respectively.

2.2.1 Data Collection

CFPB staff collected data at the Gettysburg Lab for Experimental Economics (GLEE).[15] Participants interacted through private, networked computer terminals. The experiment was conducted across 18 "sessions." Each session included 12 participants and individuals could participate in no more than one session each.

Participants earned "tokens," and at the end of each session, tokens were exchanged for dollars.[16]

In each session, participants interacted in markets. Treatments (i.e., the one price treatment, the eight sub-price treatment, and the 16 sub-price treatment) were conducted within-subjects, meaning that all participants participated in all treatments.[17] The order in which treatments were administered was randomized across sessions.

Markets in each treatment were conducted over twelve "rounds." In each round, each participant was randomly reassigned as either a buyer or seller and randomly reassigned to a new market with two total buyers and two total sellers.[18]

After each round, players viewed a summary screen (Figure 2) that detailed: (1) their total benefits, costs, and net earnings; (2) both sellers' total prices and sub-prices (if relevant); and (3) which seller each buyer purchased from. Thus, in addition to providing feedback on player's choices, this screen showed buyers what they could have earned if they had selected the other seller and also allowed sellers to estimate the amounts they could have earned had they set prices differently (e.g., if they had set their prices lower than the other seller).

---

[15] For additional details about the Gettysburg College laboratory space, see Appendix A. For additional details about the experimental procedure, see Appendix B.

[16] Tokens were exchanged at a rate of 550 tokens per dollar. In addition, they were paid a fixed amount of $7 for attending the session. In total, participants earned about $30 total, on average.

[17] However, not all these data are used in the analysis. See the subsection *Data Restrictions,* below.

[18] This randomization of participants into new markets each round allows participants to accumulate experience in the task but without accumulating experience with the other *players* in their market. In contrast, studies that keep market participants the same across market iterations may wish to examine how repeated interactions between known market participants affects market outcomes. For example, Plott (1982) discusses experiments on Bertrand duopolies by Stoeker (1980) in which participants interacted as sellers in many markets and in which participants sold at relatively high prices, on average.

FIGURE 2

An example of a seller's summary screen

Round Results

This round you were Seller 1 and earned 266 tokens.

You sold 2 objects for 233 tokens each.

Your total production cost was 200 tokens.

Your earnings are itemized in the table below.

Sellers' Prices

	SELLER 1	SELLER 2
Price	233 tokens	344 tokens

Your Earnings

	VALUE	# SOLD	TOTALS
Benefit	233	x 2	466
Cost	-100	x 2	-200
Total	133	x 2	266 tokens

Buyers' Choices

	SELLER 1	SELLER 2
Buyer 1	X	
Buyer 2	X	

In this example summary screen, the participant acted as a seller in a market in the one price treatment. The seller set their total price at 233 tokens, sold two objects, and made a profit of 266 total tokens.

2.3 Measures and Hypotheses

2.3.1 Primary Outcome Measures

Measures

We focus the analysis on three outcome measures: (1) "seller asks," the total prices set, or asked, by sellers; (2) "buyer mistakes," whether buyers bought the product with the higher total price;[19] and (3) "transaction prices," the total prices of products bought by buyers. We began the investigation with one broad hypothesis for each of these outcome measures.[20]

Hypotheses

Underlying each of the main hypotheses is the hypothesis that sellers will, on average, increase the complexity of their prices in treatments with more sub-prices. Measures of complexity and hypotheses surrounding measures of complexity are discussed in the next sub-section (Intermediate Outcome Measures).

First, we hypothesized that buyers who participate in markets for products with more sub-prices would be more likely to make mistakes. This behavior would be consistent with the assumptions of Carlin (2009), the results of Kalaycı (2015), and previous literature in information overload in which consumers who face more cognitively demanding choices make worse choices, on average (e.g., Carpenter et al., 2021). This hypothesis is summarized in hypothesis H1, below:

(H1) Buyers' frequency of making mistakes will be positively correlated with the number of sub-prices in a market.

Next, we hypothesized that sellers would set higher total prices in markets with more sub-prices. One possible explanation for this behavior would be that sellers anticipate (or react to) the relationship between buyers' mistakes and dimensional complexity hypothesized in H1 and thus

[19] Although it is possible that participants had preferences that could rationalize such choices—e.g., a preference for higher total prices, or a preference for giving other participants more money—we feel comfortable labeling these choices as "mistakes" given that they resulted in a loss of earnings for the participants who made them.

[20] Hypotheses 1-3 are predictions of correlations between outcomes and treatment. We did not have precise predictions about the magnitudes of the hypothesized effects nor how the magnitudes might vary across the domain of potential dimensional complexity (i.e., how magnitudes differed when the number of sub-prices increased from one to eight as compared when the number of sub-prices increased from eight to sixteen). In addition, although the dynamics of outcomes variables were not our primary concern, we hypothesized that each of the outcome variables would decrease as rounds progressed. That mistake frequencies would decrease is consistent with the observation that experiment participants typically improve at tasks with experience. That seller asks would decrease is consistent with other experimental evidence on Bertrand duopoly games (Fouraker and Siegel, 1963; Dufwenberg and Gneezy, 2000). We did not have hypotheses about how dynamics would be affected by treatment.

set prices higher in the hope of earning more via buyers' mistakes. This hypothesis is summarized in hypothesis H2, below:

(**H2**) Sellers' asks will be positively correlated with the number of sub-prices in a market.

Finally, we hypothesized that, because of the effects hypothesized in H1 and H2, the transaction prices for products with more sub-prices would be higher, on average. This hypothesis is summarized in hypothesis H3, below:

(H3) Transaction prices will be positively correlated with the number of sub-prices in a market.

2.3.2 Intermediate Outcome Measures

We also tracked several "intermediate measures." These variables are intermediate because they could plausibly be affected by treatment, and they could also themselves plausibly affect the primary outcome measures.

The first intermediate variables measure *realized* price complexity in the experiment's markets. To see why this is necessary, recall that treatment determines the *maximum* dimensional complexity of seller's prices. In the actual markets, sellers set their own prices, and regardless of treatment status, it is possible that sellers choose prices that are *not* complex. For example, a seller could set most of their sub-prices to zero, making the task of adding and comparing prices simpler for buyers. In addition, features of markets may emerge (but not be the direct result of any one seller's actions) that make the markets more or less complex. For example, if both sellers set most of their sub-prices to be the same as each other's sub-prices, then comparing prices would be simpler for buyers.

Comparative Complexity

We borrowed from the field of Information Theory to quantify the complexity of information in a market. Following Lurie (2004), we used "Shannon entropy" (entropy) to measure the price complexity buyers face in a market when comparing prices across sellers.[21] We calls this measure "comparative complexity" or, when discussing measures of entropy directly,

[21] Shannon entropy is a measure of the of uncertainty of a random variable. For any random variable X with support $S(X)$ and probability mass function $p(x)$, the "self-information" of an event $x \in S(X)$ is measured as $I(x) = log_2\left(\frac{1}{p(x)}\right)$. Self-information is sometimes also called the "surprisal" of x, a name that conveys that the more surprising, or less likely, x is, the greater is its value of self-information. For the random variable X, Shannon entropy is the expected value of self-information: $H(X) = \sum_{x \in S(x)} p(x)\, log_2\left(\frac{1}{p(x)}\right)$.

"comparative entropy."[22] Entropy is measured in "bits," with more bits corresponding to more complexity.[23] For each price/sub-price in a market in the experiment, if the two products have the same value, then the entropy measure corresponding to that price/sub-price is zero bits. If the two products have different values for a sub-price, then the entropy measure corresponding to that price/sub-price is one bit. Thus, for the one price treatment, comparative entropy can range between zero bits when prices are the same and one bit when prices are different; for the eight sub-price treatment, comparative entropy can range between zero bits and eight bits; and for the 16 sub-price treatment, comparative entropy can range between zero bits and 16 bits.[24]

Thus, (1) the more similar sellers' sub-prices are, the lower will be the Shannon entropy,[25] and (2) the more sub-prices there are, the greater the maximum possible Shannon entropy.[26] Shannon entropy is intuitive as a measure of comparative complexity because, intuitively, the more similar products are, the easier they are for buyers compare. Nevertheless, Shannon entropy does not encompass all intuition for what makes comparison simple or complex. For example, Shannon entropy has no notion of dominance, while a simple intuition is that if one seller's sub-prices dominate another's (i.e., their sub-prices are all the same or lower than the other seller's) then this would be a very simple price comparison for buyers.

Dimensional Complexity

Next, we use Shannon entropy to measure what we call "dimensional complexity" or, when discussing measures of entropy directly, "dimensional entropy."[27] Whereas comparative

[22] Because we apply entropy to known prices in a market, our use of entropy is like others that quantify known proportions such as the Shannon diversity index: https://en.wikipedia.org/wiki/Diversity_index#Shannon_index

[23] Although other bases are possible, this formulation of Shannon entropy, which uses log base two, is common. Under this formulation, both self-information and Shannon entropy are measured in "bits." Other bases correspond to different units.

[24] Below, we describe the Competition Experiment in which we conduct markets with up to four sellers. In these markets, the dynamics of comparative entropy are similar, with entropy for a single sub-price ranging from zero bits, if all sellers choose the same value; to two bits, if all sellers choose unique values.

[25] When self-information uses log base 2, as we specify here, self-information increases by one whenever the probability of an event decreases by half. Thus, if $p(x) = \frac{1}{2}$, then $I(x) = 1$; if $p(x) = \frac{1}{4}$, then $I(x) = 2$; and so on. Similarly, if an event occurs with certainty, then Shannon entropy achieve its minimum value of zero bits.

[26] Shannon entropy is maximized when there is a uniform likelihood of each possible event. A corollary of this is that random variables with more possible events also have a greater possible maximum entropy since, taking the uniform distribution over all events, each event becomes less likely.

[27] For dimensional entropy we made one change to the Shannon entropy formula, which was to exclude sub-prices set to zero from the calculation. We made this change to match our intuition that zeros do not increase complexity, and that the simplest possible price structure (under which the seller does not lose money) is to have one positive-valued sub-price with the remaining sub-prices set to zero.

complexity measures the complexity buyers face in a market when comparing prices across sellers, dimensional complexity measures the complexity of an individual seller's sub-prices.[28]

In the one price treatment, dimensional entropy is always zero bits because there is only a single price. In the eight sub-price treatment, dimensional entropy can range from zero bits (when a seller sets all of their sub-prices to be the same value or zero) to three bits (when a seller sets all of their sub-prices to be unique, non-zero values), and in the 16 sub-price treatment, dimensional entropy can range from zero bits to four bits.[29] For example, in the eight sub-price treatment, if a seller wanted to ask for 360 tokens, they could achieve this through a variety of combinations of sub-prices. Some of these combinations would have high dimensional entropy, and some would have low dimensional entropy. If the seller set all their sub-prices to be 45 tokens, then their dimensional entropy would be zero bits because all their sub-prices would have the same value. Similarly, if the seller set a single sub-price to be 360 tokens and set the remaining seven sub-prices to zero tokens, then their dimensional complexity would be zero bits because all their sub-prices would be the same value or zero. In contrast, if the seller set their sub-prices to be 10, 20, ..., 80, then their dimensional entropy would be three bits (the maximum possible) because all their sub-prices would be unique values (no repeats) and not equal to zero.[30]

Thus, (1) when sellers use few unique values for sub-prices, for example, by reusing values, or by setting sub-prices to zero, then this measure of Shannon entropy will be low, while if sellers use varied values, this measure of Shannon entropy will be high, and (2) the more sub-prices there are, the greater the maximum possible Shannon entropy. Shannon entropy is intuitive as a measure of dimensional complexity because, intuitively, the fewer unique values there are, the easier sub-prices are to sum-up and compare. Again, however, Shannon entropy does not encompass all intuition for what makes a seller's pricing structure simple. For example, Shannon entropy has no notion of round numbers, which, if used, would make adding sub-prices simpler.

[28] To obtain a market-level measure of dimensional entropy, the sellers' dimensional entropy measures are added together.

[29] Therefore, since there are two sellers in each market, the market-level measure of dimensional entropy ranges from zero bits to six bits in the eight sub-price treatment, and from zero bits to eight bits in the sixteen sub-price treatment.

[30] And thus, $H(X) = 8 * \frac{1}{8}\ log_2(8) = 3$. The analog for the sixteen sub-price treatment is $H(X) = 16 * \frac{1}{16}\ log_2(16) = 4$.

Seller Ask Spread

Finally, we consider the difference between sellers' total prices in each market, the "ask spread." Researchers have long observed that the difference in options' values affects their likelihood of being selected by decision-makers (Thurstone, 1927).[31]

Hypotheses

In this setting, we hypothesize that when ask spreads are low, buyers will have more difficulty in discerning which product is cheaper, and therefore buyers will be more likely to make mistakes. This would be consistent with Kalaycı (2015).

Prior to collecting data, we did not anticipate collecting or using either comparative complexity or dimensional complexity in this analysis, and therefore we did not generate hypotheses for their effects prior to running the study. However, we believe the following hypotheses are intuitive: (1) both comparative complexity and dimensional complexity will be positively correlated with the number of sub-prices in a market, (2) buyers' frequency of making mistakes will be positively correlated with both comparative complexity and dimensional complexity, and (3) sellers' total asks will be positively correlated with dimensional complexity.

2.4 Data

Each session had 12 rounds per treatment and three total treatments, and so we recorded 36 paid decisions per participant. Since there were 12 participants per session, this resulted in 432 paid decisions per session; and since there were 18 sessions, 7,776 paid decisions in total across 216 total participants.[32]

2.4.1 Demographics

After participants completed the main study, we asked them to complete a short survey. This survey contained items that we thought could plausibly correlate with participants' behavior in

[31] This logic is present in standard models of discrete choice such as multinomial logit and multinomial probit models (McFadden, 2001). It is also present in more recent models such as the Drift Diffusion Model (Ratcliff, 1978), in which decision-makers are more likely to make mistakes when the difference in product values (the "discriminability"), decreases (Clithero, 2018). This logic is also present in certain equilibrium concepts such as Quantal Response Equilibria (McKelvey and Palfrey, 1995) in which (excepting the cases of extreme parameter values) the likelihood with which decision-makers take an action is a function of the differences between the action's expected payoff and the expected payoffs of other actions (Goeree et al., 2016).

[32] Data from the first three sessions we ran were lost due to a software bug that made it impossible to download the data file at the end of the day. In addition, another session was canceled because too few people attended.

the experiment. On average, participants were about 20 years old and had completed nearly two laboratory experiments prior to this one.[33] About 56 percent of our sample was female, and English was the second language for about seven percent of the sample. These percentages are about equal to the overall Gettysburg College undergrad student population around the time of the study.[34] In addition, we measured participants' subjective numeracy (Fagerlin et al., 2007), objective numeracy (Schwartz et al., 1997), and subjective perceptions of their own willingness to take risks (Dohmen et al., 2011).[35] In all cases, average measures from our sample were greater than measures from the papers that generated the scales, reflecting differences between this and the other studies' sampling frames.

2.4.2 Data Restrictions

Analysis of seller asks and transaction prices shows a strong interaction between treatment status and the order with which the treatments are presented to consumers. In short, there is a strong correlation between the asks/transaction prices in the final round of a preceding treatment and the asks/transaction prices in the first round of the following treatment. This significantly complicates the analysis of treatment effects, and therefore, in the following analyses of these outcome variables, we restrict the sample to data collected during the first treatment administered in each session. This restriction does not affect the qualitative results of this experiment, nor does it affect whether treatment status statistically significantly affects these outcome variables. For more information see Appendix C.

[33] Possible responses for the prior experimental experience question included: '0', '1 to 3', '4 to 6', '7 to 9', '10 or more']. To estimate an average response, we used the average of each category's bounds. For the final category, we estimated 12 total experiments.

[34] About 52.6% of Gettysburg College undergraduates were female around the time of this study (https://www.usnews.com/best-colleges/gettysburg-college-3268/student-life, accessed 7/10/2017). About 6.4% of Gettysburg College undergraduates were international students around the time of the study (https://web.archive.org/web/20190909185050/https://www.gettysburg.edu/about-the-college/facts-figures/student-body?catInode=102856b9-cf11-4f5b-bb5b-05b499f4522c)

[35] Subjective numeracy is a measure of a participant's perception of their own ability to use, and preference for, numbers and probabilities. The scale we used is from Fagerlin et al. (2007) and consists of eight items on which participants rated their abilities or preferences on six-point scales. Objective numeracy is a measure of a participant's actual ability to understand and use numbers and probabilities. The scale we used consists of three items and is adapted from Schwartz et al. (1997). While subjective numeracy and objective numeracy are related, they are not equivalent. For example, Peters et al. (2019) demonstrate that high subjective numeracy benefits those with high objective numeracy but harms those with low objective numeracy. The subjective measure of risk was a single question that asked, on a five-point scale, 'How willing are you to take risks in your life, in general?' This measure has been shown to correlate strongly with actual risky behavior (see Dohmen et al., 2011).

2.5 Findings

As hypothesized, increasing the number of prices/sub-prices in a market had large and statistically significant effects on all three of the outcome measures. Table 1 displays values for the three outcome measures, averaged across all markets, separated by treatment. Table 2 reframes these data as changes from treatments with fewer sub-prices. The following sub-sections explore these main effects in greater depth.[36]

TABLE 1: MEAN OUTCOME MEASURES, BY TREATMENT

	Number of prices/sub-prices		
	1	8	16
Buyer mistakes, %	1.2 (0.3)	12.3 (0.9)	17.8 (1.1)
Seller asks, tokens	180.8 (3.0)	266.1 (4.2)	292.5 (3.3)
Transaction prices, tokens	159.4 (2.5)	236.4 (3.2)	271.9 (2.6)

Parentheses contain standard errors. For seller asks and transaction prices, n=432 for each treatment.
For buyer mistakes, observations in which seller asks are the same are dropped. Thus, n=1,210 for one price; n=1,240 for eight sub-prices; and n=1,250 for 16 sub-prices.

TABLE 2: INCREASES IN MEAN OUTCOME MEASURES, BY TREATMENT

	Number of prices/sub-prices		
	1 to 8	8 to 16	1 to 16
Buyer mistakes, %	11.1 (925%)	5.5 (45%)	16.6 (1383%)
Seller asks, tokens	85.3 (47%)	26.4 (10%)	111.7 (62%)
Transaction prices, tokens	77.0 (48%)	35.5 (15%)	112.5 (71%)

Parentheses contain percent increases.

[36] Treatment coefficients and significance levels are qualitatively similar across a range of statistical tests. The primary statistical challenge in analyzing these data is to model the data in a way that properly accounts for potential dependence between observations due to the facts that: (1) groups of individuals interacted repeatedly within a session; (2) within a session, markets within each treatment were repeated several times, and treatments were experienced in a sequence (creating the possibility of path-dependence); (3) each individual participated in many markets; and (4) multiple observations were drawn from each market (where, for example, buyers' likelihood of making errors may be correlated due to seeing the same prices). The most conservative test that avoids problems due to observation dependence is to aggregate the data up to the session-level. Despite the restrictive constraints of this approach (n=18), estimated treatment effects are similar to those indicated by raw frequencies. We also used multi-level crossed-effects logit models to attempt to conduct analyses at the individual level while accounting for the several potential levels of observation dependence. These models also yield qualitatively similar results to the raw data and to the session-level analysis but at much greater levels of statistical significance.

2.5.1 Seller Asks

Main Effects

Consistent with hypothesis (H2), average seller asks increased by about 85 tokens (47%) between markets with only one price and with eight sub-prices, and by about 26.4 (10%) between markets with eight sub-prices and markets with 16 sub-prices. Comparing one price to 16 sub-prices, average seller asks increased by more than 111 tokens, or by more than 60 percent. Each of these differences are statistically significant.[37]

Most of participants' demographic characteristics were not strongly correlated with their asks when they played the seller role. Asks were not statistically significantly related to participants' age, gender, or whether English was their first language. In contrast, on average, participants with high *subjective* numeracy (they perceived themselves to be better able to understand and use numbers and probabilities) set higher asks, on average (while objective numeracy was not statistically significantly related to seller asks).[38] In addition, participants who rated their own willingness to take risks as high were more likely to choose higher asks.[39]

The first panel of Figure 3 displays average seller asks as rounds of the game progressed, separated by treatment. The second panel presents the average low ask in each market, and the third panel presents the average high ask. Each measure is significantly lower for the one price treatment across all rounds, while the eight sub-price treatment is modestly lower than the 16 sub-price treatment in nearly all rounds. Notably, all three measures were well above the Nash equilibria of the Bertrand Duopoly game (~100 tokens) in all treatments, even in the final rounds.[40]

---

[37] Aggregating data to the session level, using the eight sub-price treatment as the comparison group, the estimated effect of the one price treatment is a reduction in seller asks of about 74 tokens, or about 30 percent ($p=0.007$); and the estimated effect of the 16 sub-price treatment is an increase in seller asks of about 37 tokens, or about 15 percent ($p=0.131$). A multi-level crossed-effects logit model at the seller decision level ($n=1,175$), controlling for certain demographic and context variables, estimates the effect of the one price treatment is a reduction in seller asks of about 31 percent ($p<0.001$) compared to the 8 sub-price treatment; and the estimated effect of the 16 sub-price treatment is an increase in seller asks of about 20 percent ($p=0.034$) compared to the 8 sub-price treatment.

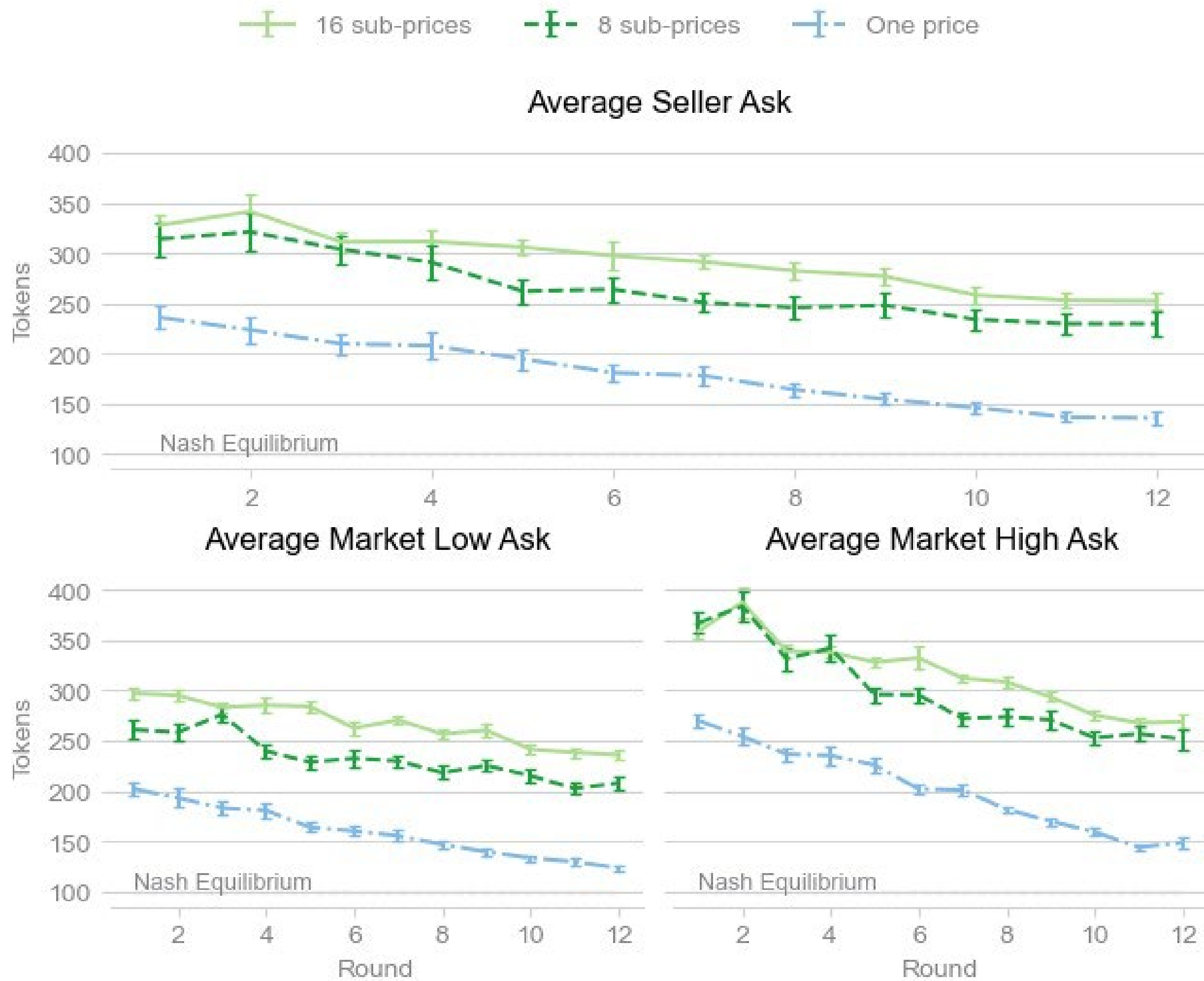
[38] Treating the measure of subjective numeracy as continuous and controlling for treatment status as well as certain context variables and several other demographic variables, a one-point increase in subjective numeracy was associated with an increase in average total ask of about 4.4 percent ($p=0.093$).

[39] Participants were asked to rate, on a scale from 1 to 5, "How willing are you to take risks in your life, in general?" Compared to participants who selected a neutral response of 3, a response of 4 is associated with an increase in average total ask of about 5.5 percent, and a response of "Very willing to take risks" (5 on the scale) is associated with an increase in average total ask of about 12.7 percent.

[40] In only two sessions, both one price treatments, did average seller asks come within ten tokens of Nash equilibrium.

FIGURE 3

Seller asks across rounds



A dashed horizontal line in each figure at 100 tokens is a symmetric Nash Equilibrium ask.

Error bars show standard errors.

Treatment appears to have affected seller asks in at least two ways. First, treatment caused a level shift in asks, which reflects sellers' expectations for a successful strategy based only on what they know about the market's structure (that is, treatment status). Considering only the first round of each session, average asks were significantly higher in the 16 sub-price treatment and significantly lower in the one price treatment, compared to the eight sub-price treatment.[41]

Second, treatment appears to have caused a change in the rate that sellers' asks declined over rounds (that is, treatment caused a change in slope). Comparing average asks in the first round to average asks in the final round, seller asks decreased more in markets with fewer prices/sub-prices.[42] In addition, relative to sellers' average asks in the preceding round, seller asks decreased more in markets with fewer sub-prices.[43]

Intermediate Outcomes

The first panel of Figure 4 shows how the dimensional entropy of sellers' prices changed as rounds progressed. Recall that dimensional complexity is determined solely by each seller (rather than as an outcome of the market), and so these data represent averages of individual choices. Since there are two sellers in each market, each market will have one seller with a lower dimensional entropy and one seller with a higher dimensional entropy. The second panel presents the average lower dimensional entropy in each market, and the third panel presents the average higher dimensional entropy. For the one price treatment, dimensional entropy was fixed at zero for all sellers/prices. For the eight sub-price treatment, the maximum dimensional entropy for a single seller was three bits and for the 16 sub-price treatment, the maximum value was four bits. A maximum dimensional entropy value indicates that all of the seller's sub-prices were unique values and greater than zero, while a value of zero bits was attained when either all of the seller's sub-prices were the same or zero. For all three measures, the 16 sub-price treatment was generally higher than the eight sub-price treatment, which was much higher than the one price treatment. Average dimensional complexity was 2.7 bits in the eight sub-price treatment and 3.15 bits in the 16 sub-price treatment. For both the eight and 16 sub-price treatments, all three measures decreased slightly as rounds progressed, on average.

---

[41] A multi-level crossed-effects logit model at the seller decision level, controlling for certain demographic and context variables and limiting data to the first round (n=106), estimates the effect of the one price treatment is a reduction in seller asks of about 19 percent (p=0.011) compared to the 8 sub-price treatment; and the estimated effect of the 16 sub-price treatment is an increase in seller asks of about 15 percent (p=0.075) compared to the 8 sub-price treatment.

[42] On average, seller asks decreased by about 100 tokens in the one price treatment, 84 tokens in the eight sub-price treatment, and 76 tokens in the 16 sub-price treatment.

[43] A multi-level crossed-effects logit model at the seller decision level (n=1,175), controlling for certain demographic and context variables, estimates that in the eight-price treatment, a one percent increase in the average ask in the preceding round leads to a 0.37% increase in average seller asks. In the one price treatment, this is reduced to about 0.33 percent ($p<0.001$); and in the 16 sub-price treatment this increases to about 0.39 percent (p=0.04).

FIGURE 4

Dimensional complexity across rounds

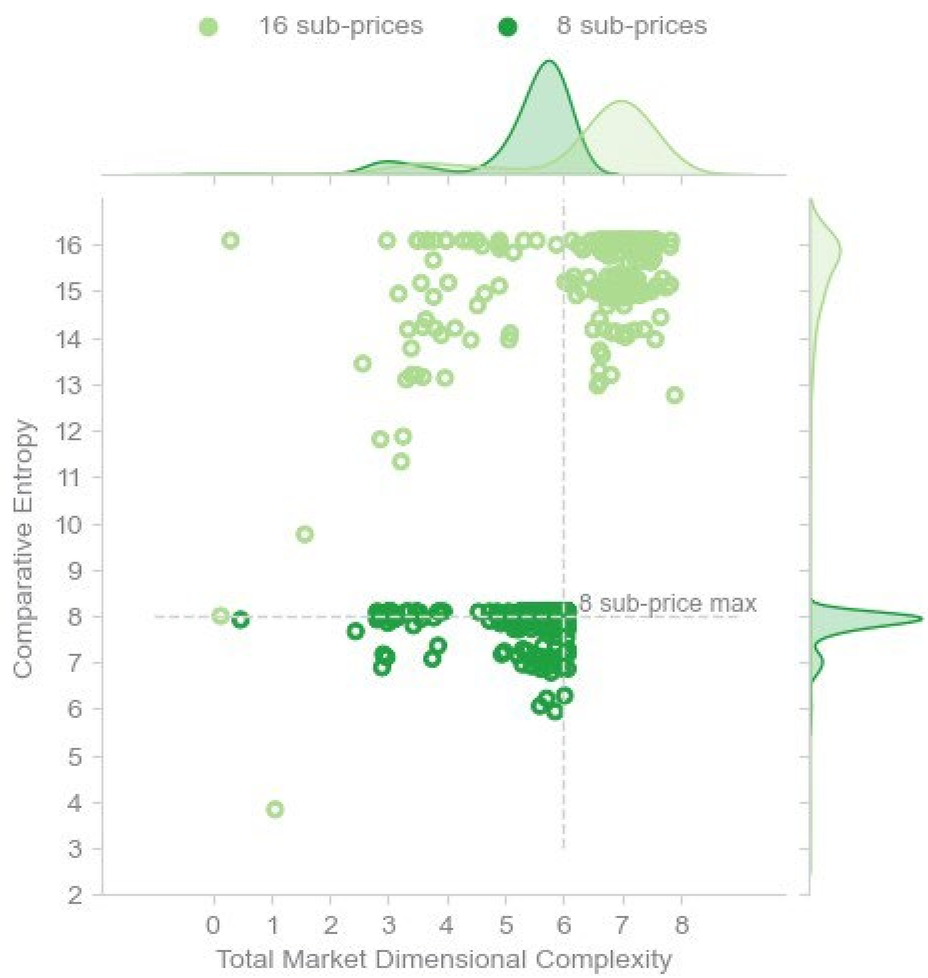
Maximal entropy values for the eight and 16 sub-price treatments are indicated by horizontal dashed lines.

Error bars show standard errors.

Figure 5 shows how comparative complexity and dimensional complexity relate to treatment status for the eight and 16 sub-price treatments. Markets from the one price treatment are not depicted because all markets with one price have a dimensional entropy of zero bits and almost all have a comparative entropy of one bit. The figure shows that *comparative* entropy is concentrated near its maximal values in each treatment. This is not surprising given that markets and roles were re-randomized before each market—sellers would have had difficulty coordinating their sub-prices, even if they wanted to. As a result, comparative entropy is almost always higher in the 16 sub-price treatment. In addition, *dimensional* entropy is also mostly concentrated near its maximal values, but, in contrast, there were several markets with dimensional entropy that was significantly lower than the maximum, indicating that some sellers made their prices simpler by repeating values or using zeros for their sub-prices.

FIGURE 5

Distributions of comparative entropy and dimensional entropy



Each point in the interior of Figure 5 represents complexity measurements from a single market. At the top of the figure is an estimated density of dimensional entropy, separated by treatment. To the right of the figure is an estimated density of comparative entropy, separated by treatment. Both entropy measures are clustered near their maximal values in markets under treatments.

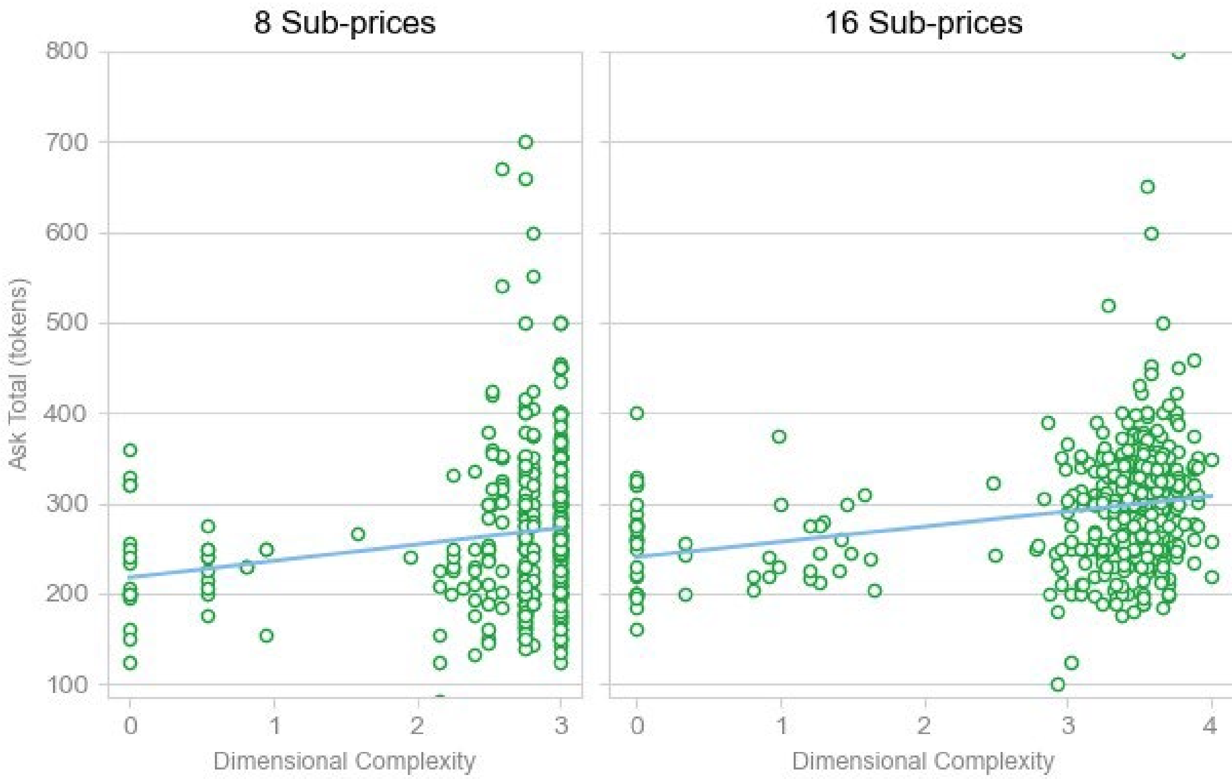
A small amount of noise has been added to each point so that multiple points with the same values can be seen.

Gray dashed lines depict the maximal values for the eight sub-price treatment.

Figure 6 shows how individual sellers' total asks relate to their dimensional complexity. These variables are both chosen by a single seller (unlike other variables that are determined at the market level), and so each datapoint reflects a decision of a single seller. Overall, there is a strong positive correlation between sellers' total ask and the dimensional complexity of their sub-prices.[44] In other words, on average, sellers who chose higher asks also chose higher complexity. At the market level, these data lead to ask spreads (not shown) being positively correlated with dimensional complexity, on average.

FIGURE 6

Seller asks and dimensional complexity



Each green circle represents one action of one seller. The solid lines use these data to predict ask totals based on dimensional entropy values.

2.5.2 Buyer Mistakes

Main Effects

Consistent with hypothesis (H1), buyer mistakes increased in frequency from 1.2 percent of buyer decisions in markets with only one price, to 12.3 percent of buyer decisions in markets

[44] Correlation coefficients were 0.39 in the eight sub-prices treatment and 0.25 in the 16 sub-price treatment.

with eight sub-prices, to 17.8 percent of buyer-decisions in markets with 16 sub-prices.[45] Comparing one price to 16 sub-prices, the frequency of buyer mistakes increased by more than 16 percentage points, or almost 1,400 percent. Each of these differences are statistically significant.[46]

Participants' demographics did not play a substantial role in their likelihood of making a mistake. Mistake likelihood was not statistically significantly related to participants' age, gender, subjective numeracy, objective numeracy, or whether English was their first language.[47]

In addition, it does not appear that buyers learned to avoid mistakes as the experiment progressed, as there was not a clear relationship between mistake rates and round number or the order in which the treatment was administered.[48] Nevertheless, it is still possible that buyers learned in the experiment but that their learning was offset by the simultaneous learning of participants in the seller role. It could be, for example, that buyers and sellers learned at similar rates which resulted in buyers' mistake rates appearing approximately static.

Intermediate Outcomes

Both dimensional complexity and comparative complexity were associated with more buyer mistakes. In addition, market ask spread was negatively correlated with buyer mistakes. That is, the closer sellers' total asking prices were to each other, the more likely it was that buyers would make mistakes. Interestingly, these three intermediate variables appear to explain most of the

---

[45] Analyses of buyer mistakes exclude markets in which both sellers set the same total price (seller asks were the same) because it is impossible for buyers to a make mistake in such cases.

[46] Aggregating data to the session level, using the eight sub-price treatment as the comparison group, the estimated effect of the one price treatment is a reduction in mistakes of about 10 percentage points ($p=0.003$); and the estimated effect of the 16 sub-price treatment is an increase in mistakes of about seven percentage points ($p=0.014$). A multi-level crossed-effects logit model at the buyer decision level ($n=3,700$), controlling for certain demographic and context variables, finds no variation at the session level. In addition, the estimated effect of the one price treatment is a reduction in mistakes of about 11 percentage points ($p<0.0001$); and the estimated effect of the 16 sub-price treatment is an increase in mistakes of about 5 percentage points ($p<0.0001$).

[47] Compared to participants who selected a neutral response (3 on a scale from 1 to 5), participants who said they were "Not at all willing to take risks" (1 on a scale from 1 to 5) were about 11 percentage points more likely to make mistakes, on average, and participants who said they were "Very willing to take risks" (5 on a scale from 1 to 5), were about 5 percentage points more likely to make a mistake, on average. Nevertheless, very few participants selected these options. Only three of the 216 participants said they were "Not at all willing to take risks" and only 12 participants said they were "Very willing to take risks."

[48] In both the one price treatment and the eight sub-price treatment, there were no statistically significant relationships between buyers' likelihood of making a mistake and either round number or the order in which the treatment was administered. In contrast, in the sixteen sub-price treatment, mistake rates *increased* at the rate of about 0.8 percentage points per round, on average, and mistakes rates were higher in the first and third treatment conditions, compared to the second.

statistical relationship between treatment and buyer-mistakes.[49] In other words, it appears the reason that the number of sub-prices affected buyers' mistakes is because more sub-prices allowed sellers to choose more complex price structures, and this caused buyers to make more mistakes, on average.[50]

Figure 7 and Figure 8 visualize the effects of the intermediate variables on buyer mistake rates. Figure 7 presents a heatmap of buyers' mistake rates, with higher rates of mistakes shown in orange and lower mistakes rates shown in blue. The figure shows that more dimensional complexity is associated with more buyer mistakes but that this relationship appears to be much stronger when ask spreads are small. In general, buyers are more likely to make mistakes when ask spreads are small and when dimensional entropy is large.[51]

Figure 8 also presents a heatmap of buyers' mistake rates but as a function of complexity. In both treatments, buyer mistakes appear to rise with each measure of complexity, and in general mistake rates are highest when both measures of complexity are high. Recall from the previous section that most markets had dimensional and comparative complexity levels near the maximum allowed, and thus most markets had combinations of these intermediate outcomes that are associated with elevated mistake rates.[52]

---

[49] Rubin (2004) demonstrates that, under certain circumstances, controlling for intermediate variables (outcomes that also depend on treatment) can lead to incorrect estimates. They note that this occurs when potential outcomes of the outcome variable (e.g., buyer mistakes) are dependent on potential outcomes of the intermediate variable (e.g., dimensional complexity) even after observing realized values of treatment and the intermediate variable. In this experiment, we assume such a relationship does not exist because sellers' actions determine intermediate variables while buyers' actions determine the outcome variable. Since, in a given market, buyers and sellers are different people, it is reasonable to assume the potential outcomes of these variables are independent.

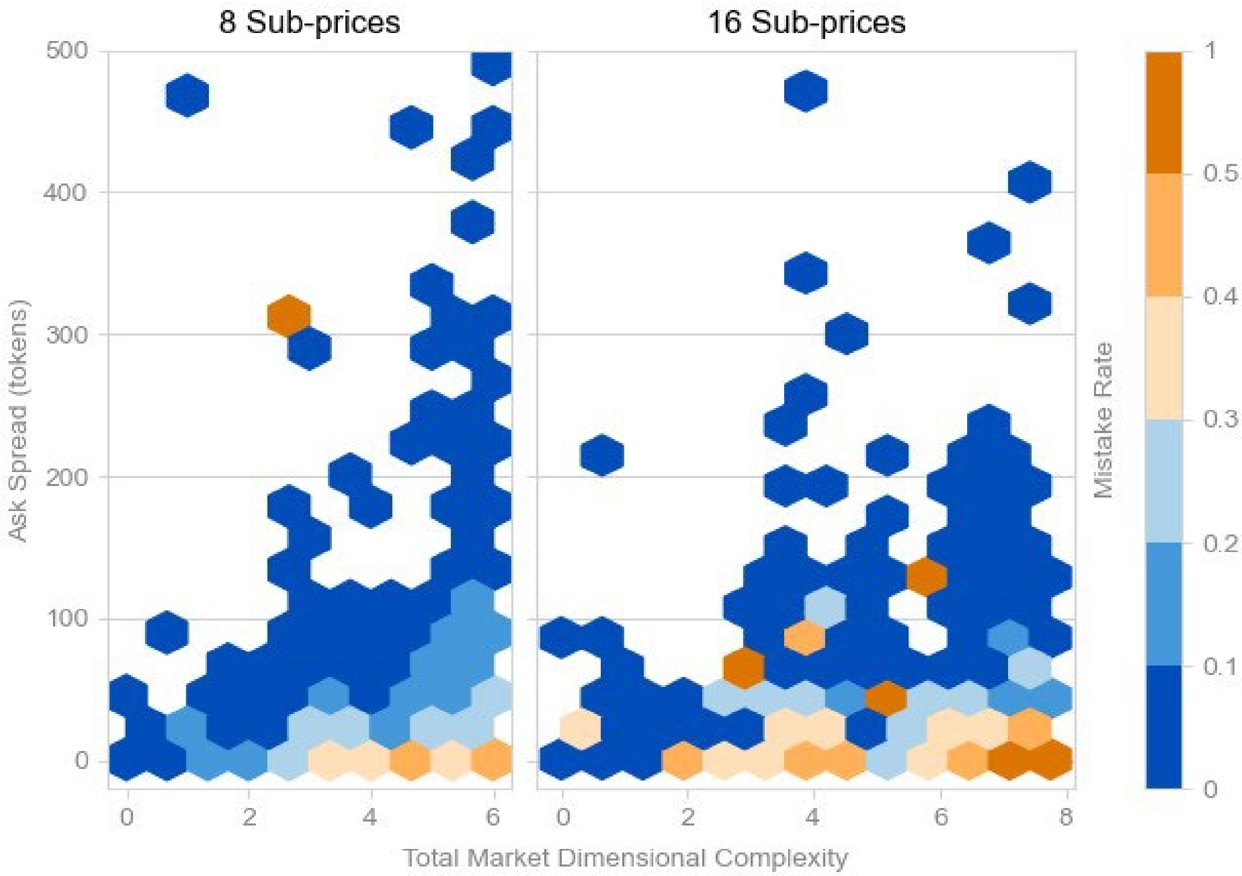
[50] Controlling for intermediate market variables and certain time, context, and demographic fixed effects, no statistically significant relationship between treatment and buyer mistake rates remain. Note that the one price treatment is highly correlated with both entropy measures. All observations in this sample have a dimensional entropy of zero bits (by definition, since there is only one number in each price, the entropy score must be zero). In addition, all observations in this sample have a comparative entropy score of one bit (recall we restrict the sample in this analysis to observations in which total prices are not equal).

[51] In a logit regression, controlling for other intermediate variables, certain time, context, and demographic fixed effects, the average effect of a 10 token decrease in ask spread is a 1.9 percentage point increase in the likelihood of buyers making a mistake ($p<0.001$). The average effect of a one bit increase in dimensional entropy is a 1.5 percentage point increase in the likelihood of buyers making a mistake ($p<0.001$).

[52] In a logit regression, controlling for other intermediate variables, certain time, context, and demographic fixed effects, the average effect of a of a one bit increase in comparative entropy is a 0.6 percentage point increase in the likelihood of buyers making a mistake ($p=0.102$). The average effect of a one bit increase in dimensional entropy is a 1.5 percentage point increase in the likelihood of buyers making a mistake ($p<0.001$).

FIGURE 7

Heatmaps of buyers' mistake rates by ask spread and dimensional complexity



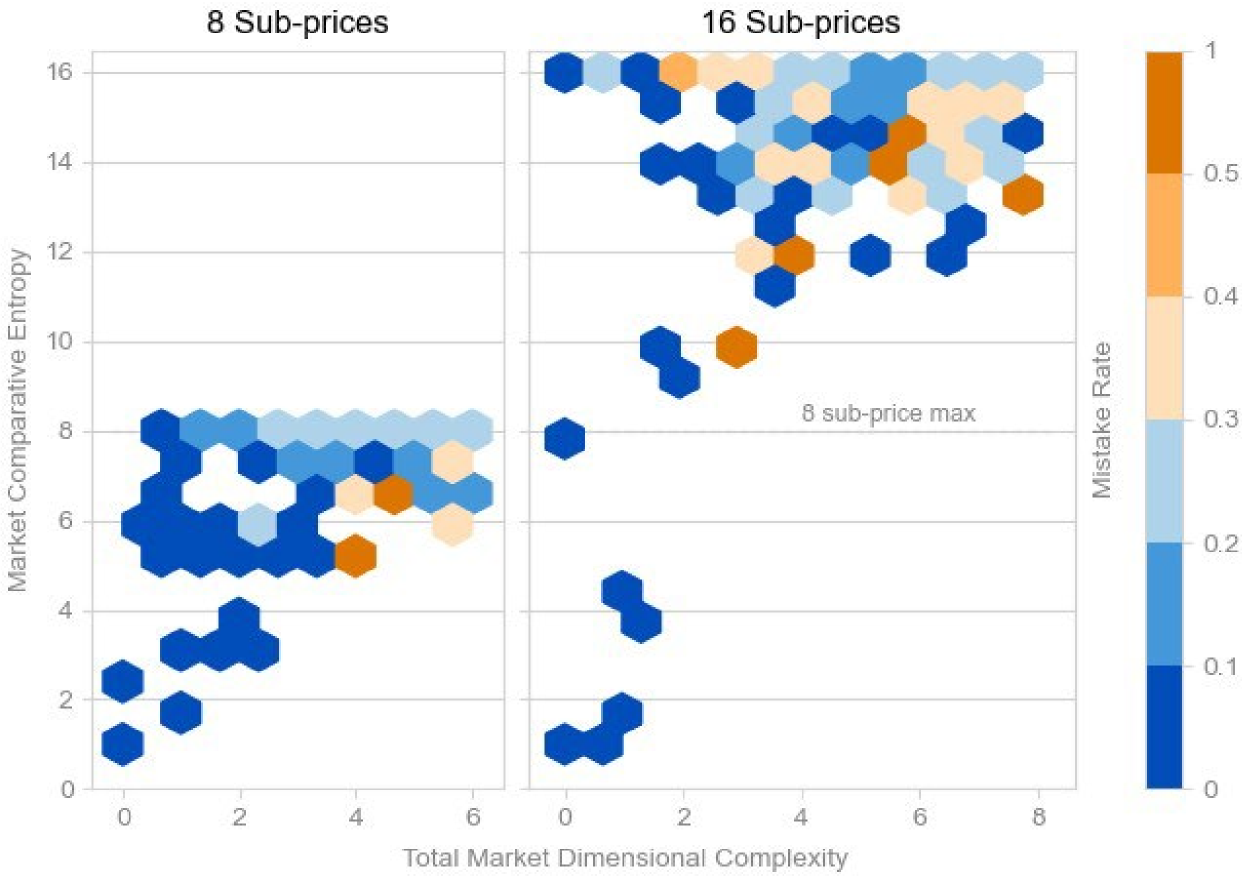
Markets with spreads greater than 500 tokens are excluded to improve figure readability.

Mistake rates are aggregated to the market level and binned with markets with similar ask spreads and total dimensional entropy.

Markets from the one price treatment is not pictured because they each had a market dimensional entropy score of zero.

FIGURE 8

Heatmaps of buyers' mistake rates by comparative entropy and dimensional entropy



Buyer mistake rates are aggregated to the market level and binned with markets with similar entropy measures.

Markets from the one price treatment is not pictured because they each had a market dimensional entropy score of zero.

2.5.3 Transaction Prices and Best Response

Transaction Prices

Consistent with hypothesis (H3), transaction prices increased by about 77 tokens (48%) between markets with only one price and markets with eight sub-prices, and by about 35.5 tokens (15%) between markets with eight sub-prices and markets with 16 sub-prices. Comparing one price markets to 16 sub-price markets, average transaction prices increased by about 113 tokens, or by more than 70 percent. Each of these differences is statistically significant.[53]

Figure 9 displays average transaction prices as rounds of the game progressed, separated by treatment. The data in this figure are only slightly higher than average market low asks (the data shown in second panel of Figure 3). This is natural, since in most cases, buyers chose the low ask in each market. On average, transaction prices did not reach the Nash Equilibria near 100 tokens.[54]

As with seller asks, treatment appears to have affected transaction prices in at least two ways. First, treatment caused a level shift in transaction prices—considering only the first round of each session, average market prices were significantly higher in the 16 sub-price treatment and significantly lower in the one price treatment than in the eight sub-price treatment.[55] Second, treatment caused a change in the rate at which transaction prices declined. Comparing average transaction prices in the first round to average transaction prices in the final round, transaction prices decreased more in markets with one price than those with multiple sub-prices.[56]

[53] Aggregating data to the session level, using the eight sub-price treatment as the comparison group, the estimated effect of the one price treatment is a reduction in transaction prices of about 65 tokens, or about 30 percent ($p=0.017$); and the estimated effect of the 16 sub-price treatment is an increase in transaction prices of about 47 tokens, or about 22 percent ($p=0.069$).

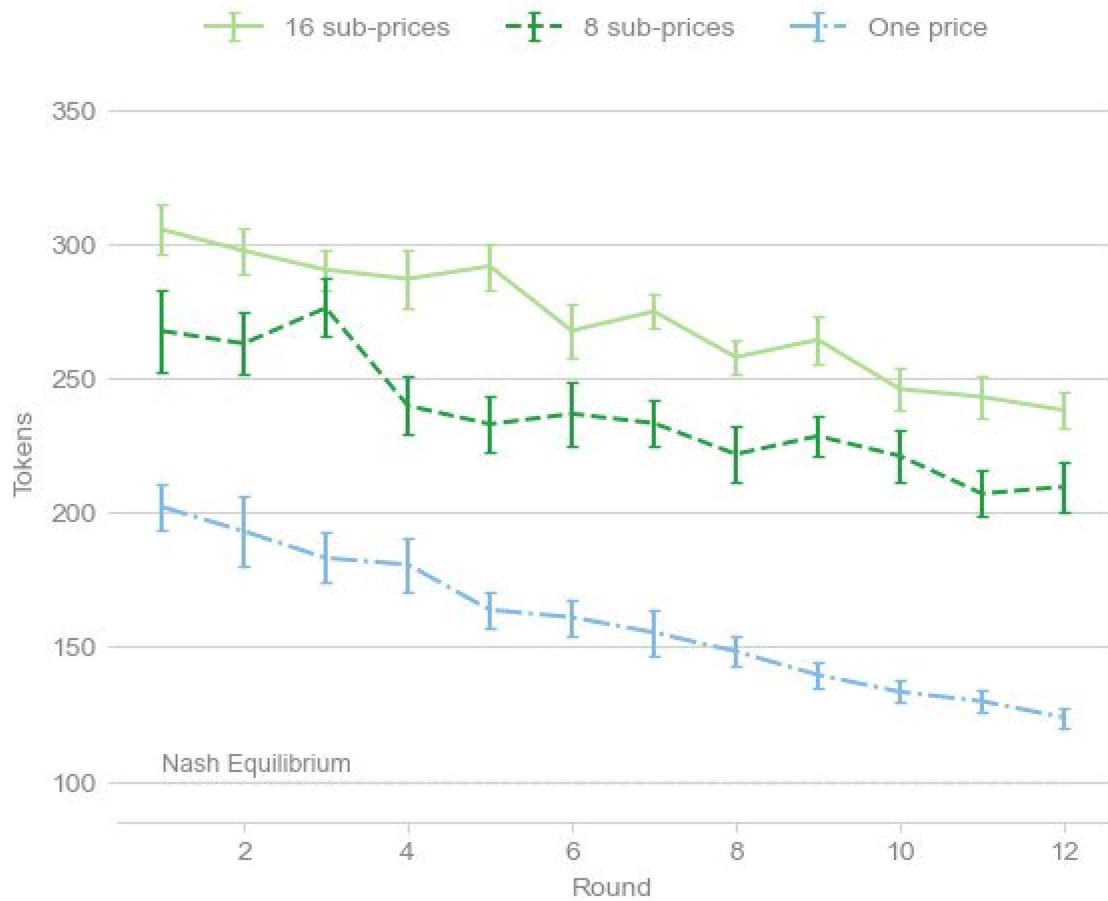
[54] In only two sessions did average transaction prices come within 10 tokens of the Nash equilibrium. Both were one price treatments.

[55] A multi-level crossed-effects logit model at the market level, controlling for certain context variables and limiting data to the first round ($n=54$), estimates the effect of the one price treatment is a reduction in transaction prices of about 19 percent ($p=0.023$) compared to the 8 sub-price treatment; and the estimated effect of the 16 sub-price treatment is an increase in transaction prices of about 24 percent ($p=0.023$) compared to the 8 sub-price treatment.

[56] On average, transaction prices decreased by about 78 tokens in the one price treatment and 62 tokens in the multiple sub-price treatments.

FIGURE 9

Average transaction prices, across rounds



The dashed horizontal line at 100 tokens is a symmetric Nash Equilibrium price.

Error bars show standard errors.

Best Response

In the previous sections, we showed that, on average, complexity leads to more buyer mistakes, higher seller asks, and high transaction prices. Asks and transaction prices trend downwards as rounds progress, but, except in the one price treatment, never approach the Nash Equilibria. Why do prices and asks never reach Nash Equilibria in the multiple sub-price treatments? Alternatively, given that more complexity results in more mistakes, why do prices and complexity not trend upwards?

To begin to address these questions, we fit a model of the likelihood that a buyer would buy from a seller, given the conditions of the market (e.g., both sellers' asks, dimensional complexity, comparative complexity, etc.) using data from the experiment.[57] This allowed us to further estimate how sellers should respond to the other seller's action, and to investigate whether adding complexity to the Bertrand Duopoly game, as we did in the experiment, changes the equilibria of the game.

Figure 10 presents a directional field graph for each treatment. Each arrow in the graph depicts a seller's "best response" if she knew what the other seller was going to do. The location of the base of an arrow represents the other seller's chosen ask total (measured on the y-axis) and dimensional complexity (measured on the x-axis). The direction in which an arrow points represents the direction of the best response. That is, the arrow points in the direction of the ask total and dimensional complexity that the responding seller should choose to earn the most tokens, in expectation, given the model of buyer behavior. The length of an arrow represents, in relative terms, how far in the indicated direction the best response is from the first seller's price and complexity.[58]

In the graph, dots without arrowheads indicate an estimated symmetric equilibrium point: a point where both sellers play the same action and neither has an incentive to change. There are only two of these, each in the one price treatment, one at an ask total of 101 tokens and the other at 102 tokens. Each of these is also a symmetric Nash Equilibrium of the standard Bertrand Duopoly game. The third equilibrium of the standard game, when both sellers ask for 100 tokens, is not an equilibrium in this model because at this price sellers make no profit, and therefore sellers would expect to earn more tokens by choosing a higher price and hoping a buyer makes a mistake.

Similarly, in the multiple sub-price treatments, when the other seller's total is near the Nash Equilibria of the Bertrand Duopoly game, the seller's best response is to *increase* their total price and to make their prices highly complex. This contrasts with the logic of the Nash Equilibria in which sellers undercut each other until their asks are at, or very near, their marginal cost of production. However, there is a clear logic to the best response pattern—when the other seller's price is low, there is very little profit to be gained by undercutting. Instead, a more profitable action is to set total price high and try to induce buyer mistakes through increased dimensional complexity.[59]

---

[57] This is a logit model like the model of buyer mistakes reported in the previous section.

[58] To be clear, arrows do not terminate at the seller's best response. This design choice was made to help clarify the figure by reducing the number crossing arrows.

[59] These patterns have much in common with Carlin's (2009) characterization of equilibrium in a similar setting. In Carlin's setting, there can exist no equilibrium in which all sellers choose the same price (and thus excludes the

FIGURE 10

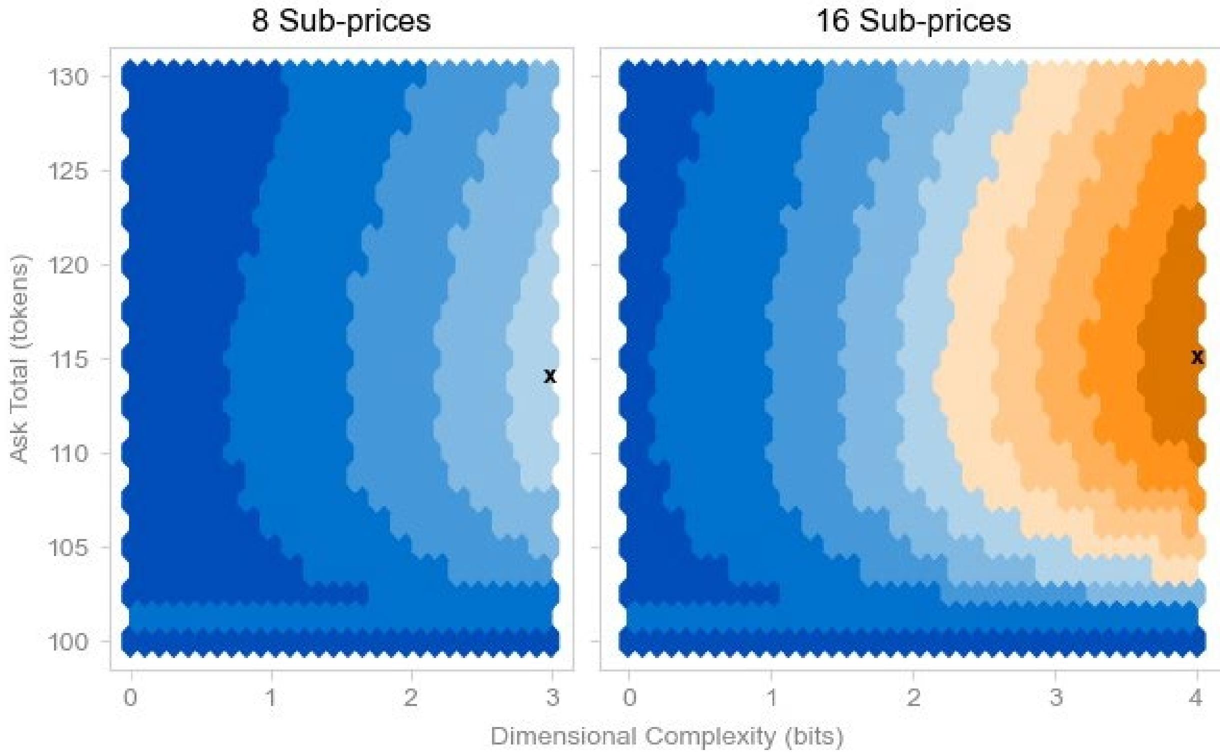
Seller's best responses given other seller's action

Each arrow in each graph depicts a seller's "best response" if she knew the ask total and dimensional complexity of the other seller. In each panel, the base of an arrow is located at the known action of the other seller. The tip an arrow points in the direction of the responding seller's best response, and the length of the arrow is proportional to the magnitude of the best response in that direction.

symmetric Nash Equilibria of the standard Bertrand game). Instead, in equilibrium, sellers play a mixed strategy in which they choose prices according to a continuous price distribution with no mass points. In addition, in addition, in Carlin's model, as in ours, sellers who choose higher prices also wish to induce mistakes by increasing complexity; and sellers who choose lower prices wish to minimize buyer mistakes by reducing complexity.

FIGURE 11

Heatmap of a seller's expected earnings



Blue hexagons indicate low expected values, with deeper hues corresponding to lower expected values. Orange hexagons indicate high expected earnings, with deeper hues indicating higher expected earnings. The black x's indicate the seller's best responses, according to the model.

Figure 11 explores a seller's expected earnings when the other seller asks for 101 tokens and chooses zero bits of dimensional complexity. This is an interesting scenario to investigate because it is one of the Nash Equilibria of the standard Bertrand Duopoly game, and it remains an equilibrium of the model in the one price treatment. Both panels of Figure 11 depict heatmaps of a seller's expected earnings for each of their possible actions. The first panel depicts expected earnings when objects have eight sub-prices, and the second panel, 16 sub-prices. Blue hexagons indicate lower expected values, with deeper hues corresponding to lower expected values. Orange hexagons indicate higher expected values, with deeper hues indicating greater expected values. Clearly, asks of 101 tokens are not a symmetric equilibrium of this game under either treatment as the seller has many options that she would expect to earn her more tokens.[60] The black x in each panel indicates the seller's best response in that treatment, according to the model. In the eight sub-price treatment, her best response is to ask for 114 tokens and choose a

[60] This result is consistent regardless of the value of the other seller's dimensional complexity.

dimensional complexity of 3 bits (the most possible). In the 16 sub-price treatment, the seller's best response is to ask for 115 tokens and choose a dimensional complexity of 4 bits (also the most possible).

In the multiple sub-price treatments, we did not find an equilibrium of the game, given the model. Instead, we found a cycle of best responses in which the best response to low prices and low complexity is higher prices and high complexity; and the best response to higher prices and higher complexity is lower prices and lower complexity; and so forth. These cycles can be seen in Figure 10.[61] Thus, if the model is correct, the dynamics of best response in the multiple sub-price treatments guarantee that ask totals (and therefore transaction prices) will never converge to the low-priced Nash Equilibria of the standard duopoly game. On the other hand, the competitive pressures at higher prices also guarantee that prices and complexity will not trend forever upwards.

[61] To clarify, we did not find any pure strategy symmetric equilibria, given the model. In addition, we did not search for mixed strategy equilibria of the game, given the model.

3. Competition Experiment

In this section, we describe Experiment 3, which, for ease of description we calls the "Competition Experiment." In addition to the manipulations conducted in the Competition Experiment, in this experiment, we also manipulated the competitive environment of markets.

One of the difficulties of identifying the effect of complexity in a market setting is that attributes of, or the composition of, the market may moderate the effects of complexity. For example, if there are many firms in a market, each competing on price, then all consumers could benefit from lower prices, and it might appear that price complexity has little effect on market outcomes.

The question we explore with this experiment is whether the amount of competition in a market ameliorates the negative effects of complexity we identified in the previous section. Previous work gives reason to be optimistic. For example, Dufwenberg and Gneezy (2000) found in single price markets that four-seller markets were more like to converge to the Nash Equilibrium than were markets with only two sellers. In addition, Kalayci (2016)[62] conducted markets in which sellers could choose to complicate their prices and found that markets with more sellers generally had lower transaction prices.[63]

Nevertheless, the results of this experiment were not a given. This experiment considers markets with greater potential dimensional complexity than any previous studies that we is aware of. In addition, increasing the number of sellers in a market also increases both the maximum possible comparative complexity and the maximum possible market-level dimensional complexity. In the Complexity Experiment, realized complexity caused buyers to make mistakes and sellers to increase asks. The question is thus whether the additional competitive forces introduced by more sellers will outweigh the effects of attendant increases in complexity.

[62] The way in which sellers could complicate prices in this experiment differs significantly from this experiment. Complications in Kalayci do potentially increase the dimensionality of object prices, but only by up to two dimensions. In addition, Kalayci's prices have the added complication that prices have unequal weights in their contribution to a total price, a complication that would not be captured by our entropy measures.

[63] Fouraker and Siegel (1963) suggest that sellers are comprised of psychological types—that the Nash Equilibrium would be obtained if at least one seller in a market is a "simple maximizing" or "rivalistic" type, that Nash Equilibrium can only be avoided if all sellers are "cooperative bargainers," and thus that the Nash Equilibrium is more likely to obtain in markets with more sellers because, as the number of sellers increases, so too does the likelihood that at least one of the sellers has a "simple maximizing" or "rivalistic" objective. On the other hand, Baye and Morgan (2004) demonstrate that markets with more sellers could be more likely to achieve lower prices, not because sellers have different motivations but because sellers have varying levels of rationality.

3.1 The Model and the Experiment

This experiment is based on the same model as the Complexity Experiment. Since the markets occasionally have four sellers, they cannot be called duopolies, and instead we simply call them Bertrand markets.

This experiment was the same as the Complexity Experiment but for two key differences.[64] First, in addition to varying the number of sub-prices in a market, we also varied the number of sellers in a market. Second, we did not test any eight sub-price treatments. Together, these changes resulted in four treatments: two sellers, one price; two sellers, 16 sub-prices; four sellers, one price; and four sellers, 16 sub-prices.[65]

3.2 Data

This experiment included 16 sessions, resulting in 192 participants, 469 markets, and 2,248 individual decisions.

3.2.1 Data Restrictions

As in the analysis of the Complexity Experiment, in our analysis of seller asks and transaction prices from this experiment, we find a strong interaction between treatment status and the order with which the treatments are presented to consumers. Therefore, in the following analyses, we restrict the sample to data collected during the first treatment administered in each session.

The software we used for this experiment contained a bug wherein participants could manually adjust their sub-prices such that their ask total was not the same as the sum of their sub-prices (these should mechanically always be the same). Since buyers only saw sub-prices, this could result in buyers paying prices that did not match the sub-prices they saw. Of the 1,344 seller asks, 11 exhibited this inconsistency. These each occurred in unique markets*, and we exclude data from these markets from the analyses*. Of the 11 excluded markets, four resulted in a profit for the seller relative to what they would have made if their total price was consistent with the

[64] There was also one implementation difference, which was that the game was served to participants via the World Wide Web from CFPB's servers rather than from the experimenter's terminal at Gettysburg College. This change did not change participant's experience.

[65] In addition, this experiment was only partially between-subjects, meaning we did not administer all treatments in all sessions. In each session, we conducted one session with one price and one session with 16 sub-prices. With this constraint, all plausible combinations number of sellers and number of prices were run and balanced across sessions. Regardless of this change, as in the analysis of Experiment 1, we are only considering data from the first treatment administered in each session. See the "Data Restrictions" sub-section for more information.

sum of their sub-prices. Each of these four instances occurred in a unique session, which means that no participant profited from this bug more than once. We conclude from these facts that no participant purposefully exploited the bug and thus we do not exclude any *sessions* from the analyses described in this report. When we do exclude all four sessions with any instances of profitable use of the bug, the results of the analyses do not qualitatively change.

3.3 Measures and Hypotheses

3.3.1 Primary Outcome Measures

The three primary outcome measures for this study are the same as before: seller asks, buyer mistakes, and transaction prices. We maintain the three primary hypotheses from above (H1, H2, and H3). In addition, we add three regarding the effects of competition.

We hypothesized that buyers would make more mistakes in markets with more sellers. Because markets with more sellers have potential both for greater dimensional complexity and greater comparative complexity, we expect that markets with more sellers will cause buyers to make more mistakes, on average.[66] This would be consistent with the results from the Complexity Experiment. This hypothesis is summarized in H4, below:

(**H4**) Buyers' frequency of making mistakes will be positively correlated with the number of sellers in a market.

Next, we hypothesized that in markets with *one price*, sellers would set lower total prices in markets with more sellers. This behavior would be consistent with previous experimental work such as Dufwenberg and Gneezy (2000). This hypothesis is summarized in H5, below:

(**H5**) In markets with one price, sellers' asks will be negatively correlated with the number of sellers in a market.

Finally, we hypothesized that, because of the effect hypothesized in H5, that the transaction prices in markets with *one price* would be lower, on average, in markets with more sellers. This hypothesis is summarized in H6, below:

[66] This contrasts with the results of Kalayci (2016) who found no effect of the number of sellers on buyer mistakes. Nevertheless, there are several differences between our studies that could lead us to reach different conclusions.

(H6) In markets with one price, transaction prices will be negatively correlated with the number of sellers in a market.

As discussed above, in 16 sub-price markets we were uncertain about how seller asks and transaction prices would correlate with the number of sellers. We therefore offered no hypotheses about the effects of additional sellers on asks or transaction prices in markets with 16 sub-prices.

3.4 Findings

3.4.1 Seller Asks

Average seller asks in each treatment are displayed in Table 3. Consistent with hypothesis (H5), average seller asks decreased by about 68 tokens (28%) when there were more sellers in one price markets.

TABLE 3: MEAN SELLER ASKS, BY TREATMENT

	One price	**16 sub-prices**	Difference	One price to 16 sub-prices*
Two Sellers	243.3 (6.3)	399.4 (9.7)	+ 156.1 (11.6)	+ 104.3 (8.4)
Four Sellers	175.4 (6.0)	238.8 (8.3)	+ 63.4 (10.2)	
Difference	- 67.9 (8.7)	-160.6 (12.8)		
Two to four sellers	- 113.9 (8.4)			- 4.5 (10.4)

Seller asks and differences are denominated in tokens. Parentheses contain standard errors.

* Values are not simple averages of the preceding column due to imbalanced sample sizes. Since two-seller markets require fewer participants, we ran 50 percent more two-seller markets than four-seller markets. In contrast, four-seller markets have 100 percent more sellers than two-seller markets. Therefore, on net, we have about 33 percent more ask data from two-seller markets than from four-seller markets (additional differences are due to unbalanced incidence of data restrictions). E.g., $(1.33 * 63.4 + 1 * 156.1)/2.33 \approx 104.3$.

We did not have clear hypotheses about how asks would change in the 16 sub-price markets—we predicted buyer errors would become more frequent which could lead sellers to choose higher asks, but that additional competitive pressure would likely encourage lower asks, and we were uncertain which force would be stronger. However, the results are clear: in markets with 16 sub-prices, more sellers led to a large decrease, about 161 tokens (40%), in average seller asks.

We also replicated the result of the Complexity Experiment in that, consistent with hypothesis (H2), average seller asks increased with the number of sub-prices in a market. Seller asks increased by about 104 tokens (51%) from about 204 tokens in markets with only one price to about 308 tokens in markets with 16 sub-prices.[67] Differences due to additional sellers and due to additional sub-prices are both statistically significant.[68]

The effects of more sellers approximately offset the effects of more sub-prices, on average. The difference between average seller asks in markets with two sellers and one price and markets with four sellers and 16 sub-prices was only about 4.5 tokens, and this difference was not statistically significant different from zero.[69]

The first panel of Figure 12 displays average seller asks as rounds of the game progressed, separated by treatment. The second panel presents the average low ask in each market, and the third panel presents the average high ask. Comparing treatments with the same number of prices/sub-prices (green lines for 16 sub-prices, blue lines for one price), average asks and average low asks are significantly lower for the four seller treatments across all rounds. Comparing treatments with the same number of sellers (darker lines for four sellers, lighter lines for two sellers), asks in one price treatments were consistently lower than asks in 16 sub-price treatments. Notably, both four seller treatments approached the Nash Equilibria of the standard Bertrand Duopoly game.[70]

---

[67] Comparing markets with two sellers to the results of the Complexity Experiment, average seller asks were higher in both the one price treatment (243 tokens compared to 181 in the Complexity Experiment) and the 16 sub-price treatment (399 tokens compared to 292 tokens in the Complexity Experiment).

[68] Aggregating data to the session level, the estimated effect of the 16 sub-price treatment is an increase in mean seller asks of about 77.2 percent (p=0.014) compared to the one price treatments and the estimated effect of the four-seller treatment is a decrease in mean seller asks of about 24.5 percent (p=0.021) compared to two-seller treatments. A multi-level crossed-effects logit model at the seller decision level (n=1,141), controlling for certain demographic and context variables, estimates the effect of the estimated effect of the 16 sub-price treatment is an increase in mean seller asks of about 53.3 percent (p<0.001) compared to the one price treatment, and the estimated effect of the four-seller treatment is a decrease in mean seller mistakes of about 39.0 percent (p<0.001).

[69] In addition, from Table 3, the effect of additional sellers appears to stronger in markets with 16 sub-prices (-160.6 tokens) than in markets with one price (-67.9 tokens). Similarly, the effect of additional sub-prices appears to be greater in markets with two sellers (+156.1 tokens) than in markets with four sellers (+63.4 tokens). However, these differences are not statistically significant: analysis at either the individual level or the session level shows an interaction between sub-prices and sellers is not statistically significant (p=0.26 and p=0.48, respectively).

[70] In the final five rounds, six sessions' average low asks came within five tokens of the Nash equilibria of the standard Bertrand Duopoly game. Four were one price, four seller sessions; and two were 16 sub-price, four seller treatments.

FIGURE 12

Seller asks across rounds

A dashed horizontal line in each figure at 100 tokens is a symmetric Nash Equilibrium ask.

Error bars show standard errors.

3.4.2 Buyer Mistakes

Mistake rates in each treatment are displayed in Table 4. Consistent with hypothesis (H4), buyer mistakes increased with the number of sellers in a market. Buyer mistakes increased by about

11.1 percentage points (103%) from about 9.9 percent of buyer decisions in markets with two sellers to about 20.1 percent of buyer decisions in markets four sellers.

TABLE 4: MEAN BUYER MISTAKE RATES, BY TREATMENT

<table>
<tr><th></th><th>One price</th><th>16 sub-prices</th><th>Difference (pp)</th><th>One price to 16 sub-prices (pp)*</th></tr>
<tr><td>Two Sellers</td><td>5.6 (1.4)</td><td>14.4 (2.1)</td><td>+ 8.8 (2.5)</td><td rowspan="2">+ 10.5 (2.5)</td></tr>
<tr><td>Four Sellers</td><td>14.6 (2.6)</td><td>27.8 (3.3)</td><td>+ 13.2 (4.2)</td></tr>
<tr><td>Difference (pp)</td><td>+ 9.0 (2.9)</td><td>+ 13.4 (4.0)</td><td></td><td></td></tr>
<tr><td>Two to four sellers (pp)</td><td colspan="2">+ 11.1 (2.5)</td><td></td><td>+ 22.2 (2.5)</td></tr>
</table>

Mistake rates are in percentages. Differences are in percentage points. Parentheses contain standard errors.

* Values are not simple averages of the preceding column due to imbalanced sample sizes. Since two-seller markets require fewer participants, we have approximately 50 percent more buyer choices from two-seller markets than from four-seller markets (additional differences are due to unbalanced incidence of data restrictions).

E.g., $(1.5 * 8.8 + 13.2)/1 * 2.5 \approx 10.5$.

We also replicated the result of the Complexity Experiment in that, consistent with hypothesis (H1), buyer mistakes increased with the number of sub-prices in a market.[71] Buyer mistakes increased by about 10.5 percentage points (114%) from about 9.2 percent of buyer decisions in markets with only one price to about 19.7 percent of buyer decisions in markets with 16 sub-prices.[72]

The combined effect of increasing both the number of sellers and the number of sub-prices was to increase buyer mistakes by about 22 percentage points (300%) from 5.6 percent in markets

[71] Comparing markets with two sellers to the results of the Complexity Experiment, mistake rates were higher in the one price treatment (5.6% compared to 1.1% in the Complexity Experiment) and lower in the 16 sub-price treatment (14.6% compared to 17.8% in the Complexity Experiment).

[72] Unlike in the Complexity Experiment, the intermediate variables do not fully mediate the effect of treatment. In each specification of the model, treatment status (or treatment status separated into number of sellers and number of sub-prices), remains highly significantly related buyer mistake rates, even when controlling for the intermediate variables. This is true even if we restrict the analysis to two-seller markets.

with one price and two sellers to 27.8 percent in markets in four sellers and 16 sub-prices.[73] Each of these differences is statistically significant.[74]

3.4.3 Transaction Prices

Average transaction prices in each treatment are displayed in Table 5. Consistent with hypothesis (H6), in markets with one price, average transaction prices decreased by about 80 tokens (39%) when there were more sellers.

TABLE 5: MEAN TRANSACTION PRICE, BY TREATMENT

<table>
<tr><th></th><th>1 price</th><th>16 sub-prices</th><th>Difference (pp)</th><th>1 to 16
sub-prices (pp)*</th></tr>
<tr><td>Two Sellers</td><td>205.7 (4.4)</td><td>348.4 (8.8)</td><td>+ 142.6 (9.8)</td><td rowspan="2">+ 102.2 (8.0)</td></tr>
<tr><td>Four Sellers</td><td>126.0 (3.5)</td><td>164.5 (5.9)</td><td>+ 38.4 (6.9)</td></tr>
<tr><td>Difference (pp)</td><td>- 79.7 (5.6)</td><td>- 183.9 (10.6)</td><td></td><td></td></tr>
<tr><td>Two to four sellers (pp)</td><td colspan="2">- 131.2 (6.7)</td><td></td><td>- 41.3 (7.4)</td></tr>
</table>

Transaction prices are in tokens. Differences are in percentage points. Parentheses contain standard errors.
* Values are not simple averages of the preceding column due to imbalanced sample sizes. Since two-seller markets require fewer participants, we have approximately 50 percent more transaction prices from two-seller markets than from four-seller markets (additional differences are due to unbalanced incidence of data restrictions).
E.g., (1.5 * 142.6 + 1 * 38.4)/2.5 ≈ 102.2.

As with seller asks, although we did not have clear hypotheses about how transaction prices would change in the 16 sub-price markets, the results are clear: in markets with 16 sub-prices, more sellers led to a large decrease of about 184 tokens (53%) in average transaction prices.

---

73 From Table 4, the effect of additional sellers appears to be stronger in markets with 16 sub-prices (+ 13.4 percentage points) than in markets with one price (+ 9.0 percentage points). Similarly, the effect of additional sub-prices appears to be greater in markets with four sellers (+ 13.2 percentage points) than in markets with two sellers (+ 8.8 percentage points). However, these differences are not statistically significant: analysis at either the individual level or the session level shows an interaction between sub-prices and sellers is not statistically significant (p=0.60 and p=0.41, respectively).

74 Aggregating data to the session level, the estimated effect of the 16 sub-price treatment is an increase in mistakes of about 11 percentage points (p=0.001) and the estimated effect of the four-seller treatment is also an increase in mistakes of about 11 percentage points (p=0.001). A multi-level crossed-effects logit model at the buyer decision level (n=874), controlling for certain demographic and context variables, finds no statistically significant variation at the session level. In addition, the estimated effect of the 16 sub-price treatment is an increase in mistakes of about 12 percentage points (p<0.001), and the estimated effect of the four-seller treatment is an increase in mistakes of about 15 percentage points (p<0.001).

We also replicated the result of the Complexity Experiment in that, consistent with hypothesis (H3), average transaction prices increased with the number of sub-prices in a market. Average transaction prices increased by about 102 tokens (60%) from about 174 tokens in markets with only one price to about 275 tokens in markets with 16 sub-prices.[75] Differences due to additional sellers and due to addition sub-prices are both statistically significant.[76]

The effects of more sellers approximately offset the effects of more sub-prices, on average. Average transaction prices decreased by about 41 tokens between markets with two sellers and one price and markets with four sellers and 16 sub-prices. The difference in average transaction prices between these sessions were not statistically significantly different from zero.[77]

Figure 13 displays average transaction prices as rounds of the game progressed, separated by treatment. The data in this figure differ only slightly from the market low asks shown in second panel of Figure 12. This is natural, since in most cases buyers chose the low ask in each market. Average transaction prices in the one price four seller treatment approached Nash Equilibrium of standard Bertrand Duopoly model by about the 6th round. On average, transaction prices in the other treatments did not reach the Nash Equilibrium of 100 tokens.[78]

[75] Comparing markets with two sellers to the results of the Complexity Experiment, average transaction prices were higher in both the one price treatment (206 tokens compared to 160 in the Complexity Experiment) and the 16 sub-price treatment (348 tokens compared to 272 tokens in the Complexity Experiment).

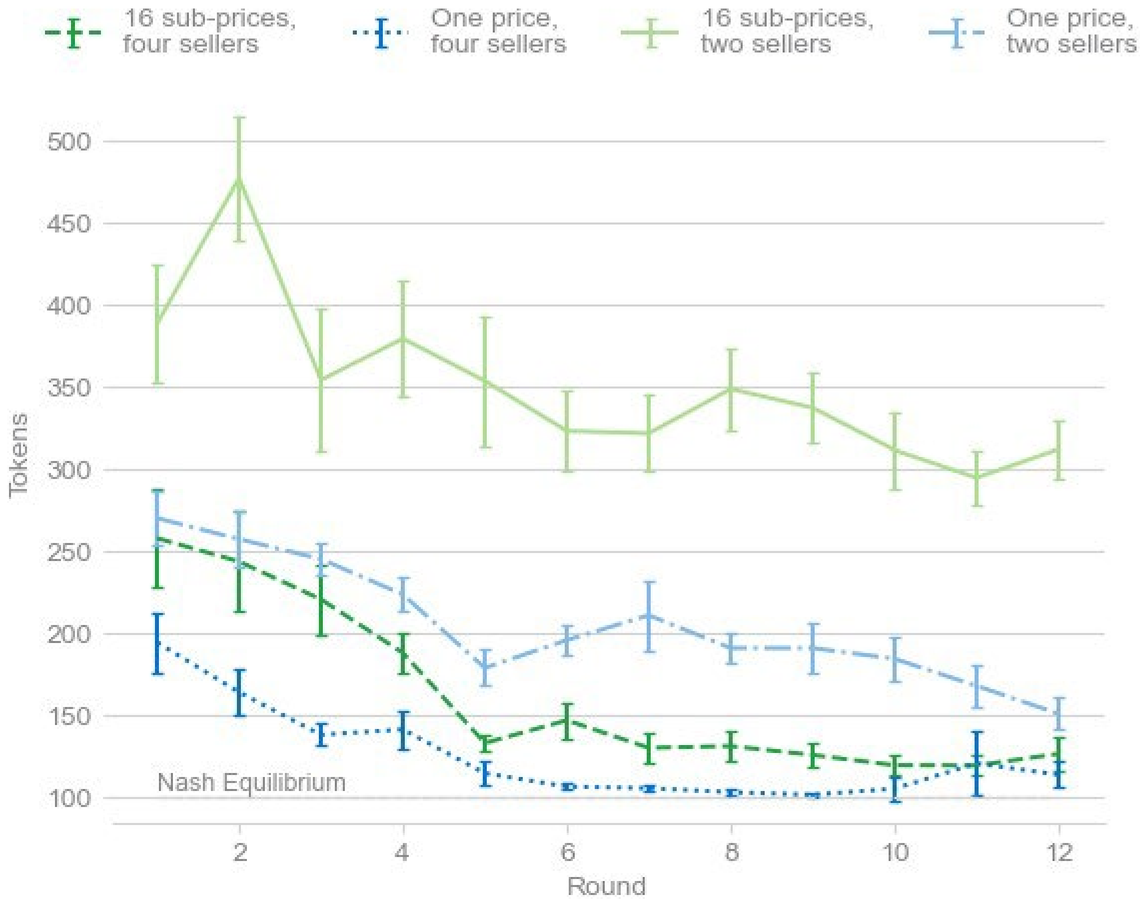
[76] Aggregating data to the session level (n=16), the estimated effect of the 16 sub-price treatment is an increase in mean transaction prices of about 43.6 percent (p=0.009) and the estimated effect of the four-seller treatment is a decrease in mean transaction prices of about 45.2 percentage (p<0.001).

[77] In addition, from Table 4, the effect of additional sellers appears to stronger in markets with 16 sub-prices (-183.9 tokens) than in markets with one price (-79.7 tokens). Similarly, the effect of additional sub-prices appears to be greater in markets with two sellers (+142.6 tokens) than in markets with four sellers (+38.4 tokens). However, these differences are not statistically significant. Analysis at the session level shows an interaction between sub-prices and sellers is not statistically significant (p=0.20).

[78] In only two sessions with treatments other than one price, four sellers did average transaction prices over the final five rounds come within 10 tokens of the Nash equilibrium. Both were 16 sub-price, four seller treatment sessions.

FIGURE 13

Average transaction prices across rounds



The dashed horizontal line at 100 tokens is a symmetric Nash Equilibrium price.

Error bars show standard errors.

4. Conclusion

In two experiments, the Consumer Financial Protection Bureau's (CFPB) Office of Research considered a variation on the standard Bertrand Duopoly model. We varied the maximum possible amount of *price* complexity in a market by varying the number of sub-prices sellers used to describe their prices. We measured *realized* complexity via measures of "comparative complexity," which quantified the dissimilarity between seller's prices, and "dimensional complexity," which quantified the dissimilarity within a single seller's sub-prices.

We find that price complexity is detrimental to consumers. On average, additional complexity (as measured by the maximum possible amount or by the realized amount) led sellers to set higher prices, increased the likelihood that consumers made mistakes, and increased transaction prices.

Sellers in both experiments generally used most of the potential complexity available to them. In the first experiment (the "Complexity Experiment"), in the eight sub-price treatment, dimensional complexity averaged nearly 90 percent of its potential (2.7 of 3.0 bits), and in the 16 sub-price treatment, dimensional complexity averaged nearly 80 percent of its potential (3.2 of 4.0 bits). In addition, sellers who chose higher dimensional complexity also chose higher total prices (r=0.30), suggesting that sellers understood the effects of complexity on buyer decision-making.

Using data from the Complexity Experiment, we fit models of buyer and seller choice. We find that, in the one price treatment, the logic of the Bertrand Duopoly mostly still holds: generally, a seller's best response is to slightly undercut the other seller's price. As in the standard Bertrand Duopoly model, in our one price markets, it remains an equilibrium for both sellers to set their prices just above their marginal cost of production. In contrast, in the multiple sub-price markets, our model shows that the logic changes substantially near the equilibria of the standard Bertrand Duopoly game. When one seller's ask is low, then the other seller's best response is not to compete on price, but to *increase* their total price while also choosing a high dimensional complexity, anticipating that, in expectation, enough buyers will make mistakes to make this a more profitable strategy.

These differences likely explain why our data generally show lower starting asks in the one price treatment as well as steeper ask declines as the experiment progresses across rounds. In addition, it is only in one price treatments that any session approaches the Nash Equilibria of the Bertrand game.

In a second experiment (the "Competition Experiment"), we found that increasing market competition ameliorated, but did not eliminate, the negative effects of complexity. Doubling the number of sellers in markets from two to four decreased average seller asks by about 36 percent, on average. These changes also had the effect of narrowing the difference in markets with more complexity compared to those with less. With two sellers, seller asks were about 64 percent higher in markets with more complexity. In contrast, with four sellers, seller asks were only about 36 percent higher in markets with more complexity. Clearly, complexity continued to affect seller asks, even in competitive markets, but its effects were substantially reduced.

Buyer and seller dynamics are reflected in market transaction prices. In these experiments, transaction prices were higher in markets with greater complexity. In the Complexity Experiment, in markets with low competition, transaction prices increased by about 70 percent in the most complex markets compared to the least; and in markets with higher competition, transaction prices increased by about 30 percent in the most complex markets compared to the least. Thus, complexity remained detrimental to consumers even in competitive markets, but its effects were substantially reduced.

Since transaction prices were paid by buyers to sellers, differences in transaction prices can be interpreted as transfers in economic well-being from buyers to sellers. For example, reinterpreting the results of the Complexity Experiment, average buyers were about 18 percent worse off in the most complex markets, while the average seller's profits increased by almost 200 percent.

Part of the value of laboratory experiments is that we can simplify decision-making environments and focus on the elements most likely to answer our research questions. In these experiments, we created simple markets and manipulated the complexity of prices and the level of competition so that we could evaluate the effects of price complexity and competition on market outcomes. These are not questions that could be easily answered in the field, and so the laboratory was a useful tool.

Nonetheless, prudence must be taken when generalizing results from laboratory experiments to the real world. The real world is, inevitably, more complicated. For example, price complexity in the real world rarely exists independently of product features; and considering features could affect consumers' choices, and, therefore, alter the insights gained from these experiments.

To be confident that laboratory results will generalize, more work must be done. Additional experiments can explore the robustness of laboratory results under different decision-making conditions, theoretical modeling can incorporate empirical insights and suggest new research questions, and, ultimately, field tests can reveal whether laboratory insights and theoretical predictions hold in real world environments.

References

Baye, M. R., & Morgan, J. (2004). Price Dispersion in the Lab and on the Internet: Theory and Evidence. *RAND Journal of Economics*, 449-466.

Bayer, R. C., Wu, H., & Chan, M. (2014). Explaining Price Dispersion and Dynamics in Laboratory Bertrand Markets. *Pacific Economic Review*, *19*(3), 278-295.

Bertrand, J. (1883). Théorie mathématique de la richesse sociale, par Léon Walras, professeur d'économie politique à l'académie de Lausanne, Lausanne, 1883. Recherches sur les principes mathématiques de la théorie de richesses, par Augustin Cournot, Paris, 1838. *Journal des Savants*, *1883*, 499-508.

Brown, A. L., Camerer, C. F., & Lovallo, D. (2012). To review or not to review? Limited strategic thinking at the movie box office. *American Economic Journal: Microeconomics*, *4*(2), 1-26.

Carlin, B. I. (2009). Strategic price complexity in retail financial markets. *Journal of financial Economics*, *91*(3), 278-287.

Carpenter, J., Huet-Vaughn, E., Matthews, P. H., Robbett, A., Beckett, D., & Jamison, J. (2021). Choice architecture to improve financial decision making. *Review of Economics and Statistics*, *103*(1), 102-118.

Chen, D.L., Schonger, M., Wickens, C., 2016. oTree - An open-source platform for laboratory, online and field experiments. Journal of Behavioral and Experimental Finance, vol 9: 88-97

Chernev, A., Böckenholt, U., & Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, *25*(2), 333-358.

Clithero, J. A. (2018). Response times in economics: Looking through the lens of sequential sampling models. *Journal of Economic Psychology*, *69*, 61-86.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association*, *9*(3), 522-550.

Dufwenberg, M., & Gneezy, U. (2000). Price competition and market concentration: an experimental study. *international Journal of industrial Organization*, *18*(1), 7-22.

Fagerlin, A., B. J. Zikmund-Fisher, P. A. Ubel, A. Jankovic, H. A. Derry, and D. M. Smith (2007), 'Measuring numeracy without a math test: Development of the subjective numeracy scale', Medical Decision Making, 27, 672–680. doi: 10.1177/0272989X07304449

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *The Quarterly Journal of Economics*, *121*(2), 505-540.

Goeree, J. K., Holt, C. A., & Palfrey, T. R. (2016). Quantal response equilibrium. In *Quantal Response Equilibrium*. Princeton University Press.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, *24*(3), 461-483.

Kalaycı, K., & Potters, J. (2011). Buyer confusion and market prices. *International Journal of Industrial Organization*, *29*(1), 14-22.

Kalaycı, K. (2015). Price complexity and buyer confusion in markets. *Journal of Economic Behavior & Organization*, *111*, 154-168.

Kalaycı, K. (2016). Confusopoly: competition and obfuscation in markets. *Experimental Economics*, *19*, 299-316.

López-Acevedo, G. (1997). Quantal response equilibria for posted offer-markets. *Estudios Económicos*, 95-131.

Lurie, N. H. (2004). Decision making in information-rich environments: The role of information structure. Journal of consumer research, 30(4), 473-486.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, *10*(1), 6-38.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 380-391.

Morwitz, V. G., Greenleaf, E. A., & Johnson, E. J. (1998). Divide and prosper: consumers' reactions to partitioned prices. *Journal of marketing research*, *35*(4), 453-463.

Peters, E., Shoots-Reinhard, B., Tompkins, M. K., Schley, D., Meilleur, L., Sinayev, A., ... & Crocker, J. (2017). Improving numeracy through values affirmation enhances decision and STEM outcomes. *PLoS one*, *12*(7), e0180674.

Plott, C. R. (1982). Industrial Organization Theory And Experimental Economics. *Journal of Economic Literature*, *20*(4), 1485-1527.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Rubin, D. B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics*, *31*(2), 161-170.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. Annals of internal medicine, 127(11), 966-972.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, *5*(1), 3-55.

Sheridan, S. L., & Pignone, M. (2002). Numeracy and the medical student's ability to interpret data. *Effective clinical practice: ECP*, *5*(1), 35-40.

Stoecker, R. (1980). *Experimentelle Untersuchung des Entscheidungsverhaltens im Bertrand-Oligopol*.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, *34*(4), 273.

Viscusi, W. K. (1978). A note on" lemons" markets with quality certification. *The Bell Journal of Economics*, 277-279.

Voester, J., Ivens, B., & Leischnig, A. (2017). Partitioned pricing: review of the literature and directions for further research. *Review of Managerial Science*, *11*, 879-931.

White, S. M., Sussman, A. B., Beckett, D., & Hong, S. (2023). When Shrouding Prices Signals Transparency: A Preference for Costly Disclosures. *Available at SSRN 4479218*.

APPENDIX A: THE GETTYSBURG COLLEGE LAB

The Gettysburg Lab for Experimental Economics (GLEE) space is a computer lab containing 27 networked computer terminals, each separated by physical dividers that prevent participants from viewing other participants' computer monitors.[79] The lab has large television monitors at the front of the room which can be seen clearly from all computer terminals and can be used to display instructions or other public information. Experimenters can observe and control experiments from an adjacent room that contains additional computer terminals and an observation window into the computer lab. A second adjacent room contains lockers for participants' personal belongings and serves as both an entry and exit to the lab for participants.

The software for the Complexity Experiment and the Competition Experiment was created using oTree version 1.2.5, a software platform built atop tools for designing webpages.[80] The Complexity Experiment was served by a computer in the observation room, while the Competition Experiment was served from CFPB via the internet. Participant's terminals ran Microsoft Windows and accessed the experiment via the Google Chrome web-browser.

---

[79] Further information about GLEE can be found on the GLEE website: https://www.gettysburg.edu/academic-programs/economics/glee/

[80] See https://otree.readthedocs.io/en/latest/index.html

APPENDIX B: THE EXPERIMENTAL PROCEDURE

Sessions of each of the experiments proceeded in essentially the same way. Prior to participants' arrival, experimenters logged-in to the terminals, navigated the web-browser to the experiment's URL, and then entered the browser into full-screen mode to hide the computer's navigation interfaces (e.g., the start menu and the task bar) as well as the browser's navigation and menu interfaces (e.g., the address bar and the back button). When participants arrived, they were instructed to not attempt to leave the browser window or to attempt to navigate away from the experiment's webpage. Because the experiment was managed by a central server, if a participant did somehow manage to navigate away from the experiment's URL, then experimenters could return the participant to the appropriate part of the experiment simply by navigating the participant's browser back to the experiment's URL.

When participants arrived through the room with the lockers (see Appendix A), they were asked by an experimenter to sign Gettysburg College's informed consent document.[81] They were then assigned a locker where they were asked to put all their personal items, including their mobile phones. Participants were then seated at a computer terminal. To ensure privacy, we left vacant every other computer terminal. In addition, to prevent attempts to communicate between participants, participants who arrived together were placed far apart from each other in the lab. As participants were seated, the monitors at the front of the lab displayed instructions asking participants to sit quietly and await further instructions.

Once participants were seated, an experimenter explained that experiment instructions and all interactions would take place via their computers, asked participants to raise their hands if they had questions during the experiment, and then explained the procedure for how participants would be paid and dismissed at the end of the experiment. The experimenter then answered participant questions and began the experiment.

All questions raised before or during the experiment were answered by the experimenter. Clarifying questions were answered for all to hear so that all participants had the same information.

Participants read instructions at their own pace. The instructions included several interactive "quiz" questions to ensure participants comprehended the instructions. Regardless of their responses to these questions, participants received feedback about which answer was correct and why.

[81] CFPB's privacy statement was also provided but as part of the game's instructions.

After the initial instructions were completed, instructions were provided about the specifics of the first treatment, these were followed by two unpaid practice rounds of the treatment-specific game, and these were followed by 12 paid rounds of the game. This process was then repeated for the remaining two treatment arms.

On each non-instruction experiment screen, participants saw a timer that counted down from 60 seconds. It was explained to participants that the timer had no direct effects and that nothing would happen if it reached zero. Nevertheless, participants were asked to make their choices approximately within the sixty seconds displayed by the timer.[82]

After all rounds for all treatments were complete, participants completed a survey comprised of psychometric scales and demographic questions. Finally, after participants completed the survey, experimenters called participants one-by-one to the adjacent monitoring room, where they were paid their earnings in cash, in private. Participants were provided with a simple word search game to play while they awaited their turn to be paid.

---

[82] The timer was implemented after pilot tests revealed that a small proportion of participants took up to several minutes with each choice. This behavior was not observed again after the timer and instructions asking for timely decision-making were implemented.

APPENDIX C: TREATMENT BLOCKS

This appendix discusses the evidence for removing treatment blocks after the first. Data presented in this Appendix are from the Complexity Experiment.

In the second treatment block, first round asks were highly correlated with final asks of preceding treatment. In treatment blocks after the first, this had the effect of increasing one price treatment prices relative to the others and decreasing 16 sub-price treatments relative to the others. Table 6 presents average seller asks in the first round of the second treatment block. Average asks are higher in the one price treatment when it is preceded by 16 sub-prices than when it is preceded by eight sub-prices. Similarly, average asks are lower in the 16 sub-price treatment when it is preceded by the one price treatment than when it is preceded by the eight sub-price treatment. Both differences are statistically significant at the five percent level. There is not statistically significant difference between eight sub-price treatments by which other treatment preceded it.[83]

TABLE 6: AVERAGE FIRST ROUND SELLER ASKS IN THE SECOND BLOCK, BY PRECEDING TREATMENT

	Treatment		
Preceding Treatment	**One Price**	**8 sub-prices**	**16 sub-prices**
One Price	--	236.4 (25.8)	259.2 (28.7)
8 sub-prices	220.1 (10.9)	--	342.4 (24.8)
16 sub-prices	267.9 (15.7)	221.5 (9.1)	--

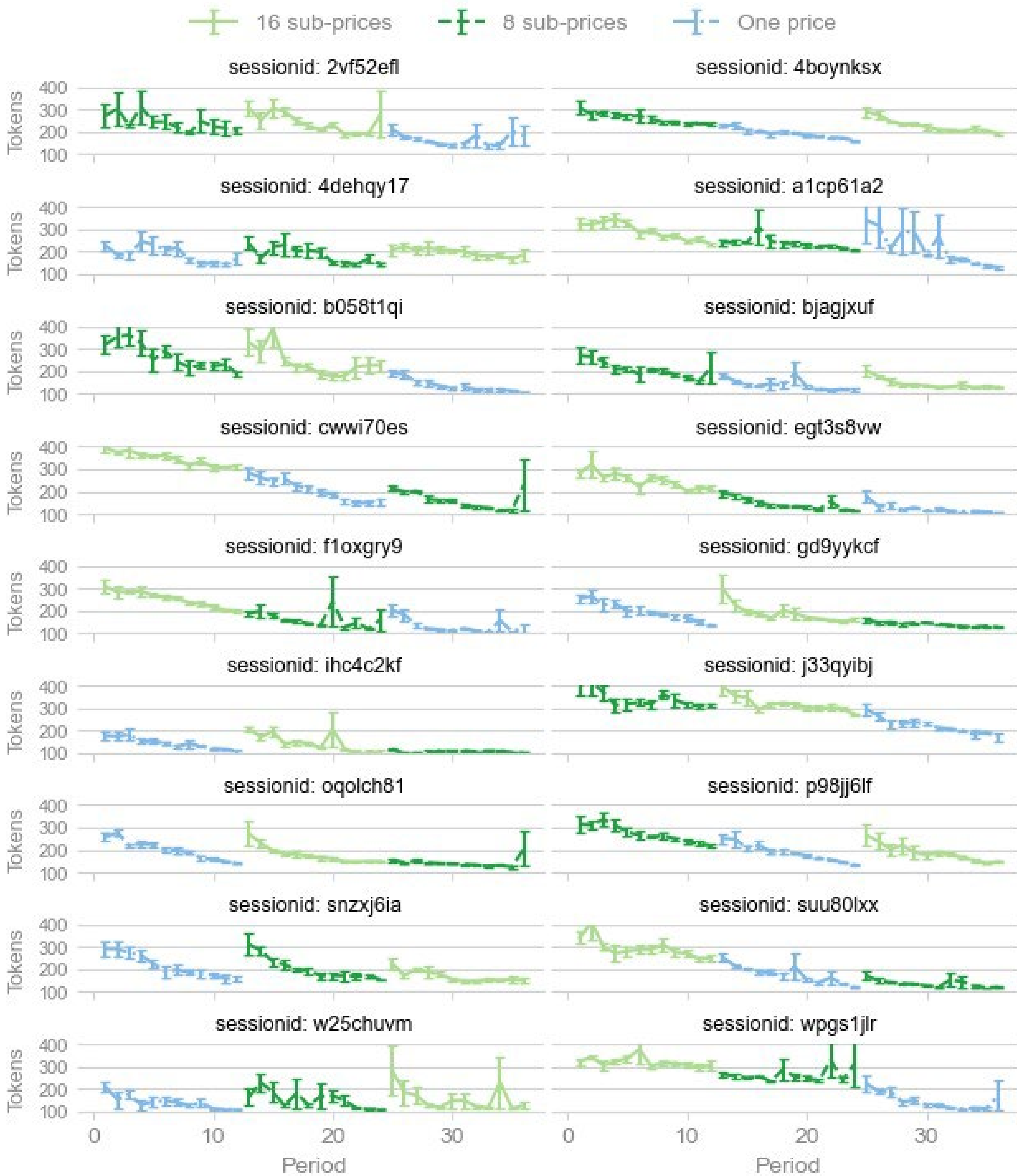
Average asks are in tokens. Parentheses contain standard errors.

Figure 14 show average asks in each of the 18 sessions we ran across all treatment blocks and rounds. As can be seen, average asks in the final round of a preceding treatment block are strong predictors of average asks in the first round of the next treatment block. It also appears that discrete jumps upward may occur between treatment blocks when a more complex treatment follows a less complex treatment.

[83] T-tests, two-sided, unequal variance. One price, p=0.021; 8 sub-prices: p=0.5913; 16 sub-prices: p=0.035.

FIGURE 14

Average seller asks by session



Error bars show standard errors.