ISSN (Print) 0473-453X
Discussion Paper No. 1131
ISSN (Online) 2435-0982

MARKET CONCENTRATION AND INCENTIVES TO COLLUDE IN COURNOT OLIGOPOLY EXPERIMENTS

Nobuyuki Hanaki
Aidas Masiliūnas

April 2021

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Market Concentration and Incentives to Collude in Cournot Oligopoly Experiments*

Nobuyuki Hanaki[†] Aidas Masiliūnas[‡]

April 19, 2021

Abstract

Multiple Cournot oligopoly experiments found more collusive behavior in markets with fewer firms (Huck et al., 2004; Horstmann et al., 2018). This result could be explained by a higher difficulty to coordinate or by lower incentives to collude in markets with more firms. We show that the Quantal Response Equilibrium can explain how the change in incentives alone could result in more collusive output in smaller markets. We propose a new method to manipulate the group size while keeping constant the locations of key outcomes, payoffs at these outcomes and the incentives to collude. Experiments using this normalized payoff function find that the number of firms has no direct effect on the average output or profit. We conclude that higher rates of aggregate collusion in markets with fewer firms are driven by the changes in incentives or focality rather than purely the number of firms. These findings imply that antitrust policies aimed at preventing collusion should focus on incentives rather than on the market concentration.

Keywords: experiment, oligopoly, collusion, group size, Quantal Response Equilibrium

JEL classification: C72 C91 D43 D83

*We would like to thank participants at the Second Wuhan Cherry Blossom Workshop in Experimental Economics, 2020 ESA Global Virtual Conference and Vilnius Winter Meeting 2020. We gratefully acknowledge financial support from the Aix-Marseille School of Economics, Joint Usage/Research Center at ISER, Osaka University, Japan Society for the Promotion of Science (18K19954, 20H05631), and the French government-managed l'Agence Nationale de la Recherche under Investissements d'Avenir UCA^{JEDI} (ANR-15-IDEX-01).

[†]Institute of Social and Economic Research, Osaka University, 6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan. Email: nobuyuki.hanaki@iser.osaka-u.ac.jp

[‡]Global Asia Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076. Email: aidas.masiliunas@gmail.com

1 Introduction

Cournot oligopoly experiments consistently find more tacit collusion in markets with fewer competing firms, a result termed the "number effect" (see Horstmann et al., 2018, for a recent overview and meta-analysis). The mechanism behind this effect has never been studied. If it is driven by an increased difficulty to coordinate with many participants, collusion could be prevented by regulating the market concentration. On the other hand, the number effect could be driven by changes in the incentives to collude, which would suggest focusing on the gains from collusion instead of market concentration. We show that both the Friedman index of collusion and the Quantal Response Equilibrium (QRE) predict more collusive behaviour in smaller markets in the standard setting considered in previous experiments. We then introduce a method to manipulate the market size while keeping constant the incentives to collude and the locations of the collusive, competitive, and Nash equilibrium outcomes. Our experiments replicate the standard result of more collusion in smaller markets using the standard treatments, but find no number effect when the incentives are held constant. We conclude that if the incentives are held constant, the number of competing firms does not affect the collusion rates.

Understanding how the number of interacting parties affects collusion rates is important for multiple reasons. First, it is important to know which behavioral regularities found in small scale experiments are likely to generalize to real life situations with a large number of participants. Previous research identified systematic differences between small and large groups in various games (Diederich et al., 2016, Choi et al., 2020, Hommes et al., 2020). We extend this line of research by studying both the consequences and the mechanism through which the group size affects choices. Second, it is useful to understand which factors influence collusion rates to identify the potentially collusive markets and assess the consequences of antitrust policies (Levenstein and Suslow, 2006). Historically, regulators focused on market concentration, measured either by the number of significant competitors in the market or by the Herfindahl-Hirschman Index.[1] Recently, the focus shifted towards the economic factors that affect the incentives to collude (Shapiro, 2010) and it has been debated whether an indicator of upward pricing pressure should replace the concentration-based methods (Farrell and Shapiro, 2010; Jaffe and Weyl, 2013). Whether mergers should be screened based on market concentration or on the incentives to increase prices depends on the nature of the number effect. We show that the results from the previous Cournot experiments (e.g. Huck et al., 2004) could be driven either by the reduced market size or by the accompanying increase in the incentives to collude. When we manipulate only the market size, aggregate collusion rates do not change, suggesting that market concentration alone does not have a large effect on the collusion rates and regulators might be right to focus on the incentives rather than market concentration.

We find that the market size has no direct effect on the average output or profits, but it affects the distribution of output. Specifically, we find higher variance in smaller markets, therefore more choices are classified as collusive – but also as competitive. This result could be explained by the difference in how the market size affects the variance of the payoff-relevant statistic, either the mean or the sum of output. The variance of average output produced by other firms decreases in larger markets, therefore when payoffs are a function of the average output, extremely high or low output is unlikely to maximize profits. The variance of total output does not decrease in larger markets, therefore this effect does not apply to standard Cournot treatments. We formalize the intuition of the variance effect using a modification of the QRE, in which choice probabilities are derived from the likelihood that each action is the best response. We show that this solution concept fits data better than QRE and can explain why variance is higher in smaller markets.

[1] See Davies et al. (2011), Farrell and Shapiro (1990) and U.S. Department of Justice & Federal Trade Commission, Horizontal Merger Guidelines (2010), available at http://www.justice.gov/atr/public/guidelines/hmg-2010.pdf

Our study extends the findings from the previous literature that studied the relationship between collusion and market size in Cournot oligopoly experiments. Fouraker and Siegel (1963) found slightly more collusive behavior in Cournot duopolies than in triopolies. Huck et al. (2004) found that average output is more collusive and more markets are classified as collusive in duopolies than quadropolies. Roux and Thöni (2015) replicated these results in baseline treatments without punishment. Waichman et al. (2014) found a higher frequency of collusion counts in duopolies than in triopolies when communication was not possible, although the effect was not significant in the manager sample. Similarly, Fonseca et al. (2018) found an increase in the collusiveness of output when the number of firms decreases from six to four and from four to two, in treatments without communication. In quadropolies without a forward market, output was more competitive than the equilibrium prediction, but more collusive than equilibrium in duopolies (Le Coq and Orzen, 2006). Van Koten and Ortmann (2013) also ran baseline treatments without a forward market and replicated the result of competitive output in quadropolies but output in triopolies and duopolies was not significantly different from the equilibrium prediction. Horstmann et al. (2018) found that duopolies were more collusive than triopolies, which in turn were more collusive than quadropolies, as measured using two collusion indexes. Horstmann et al. (2018) also performed a meta-analysis of previously published results and found higher rates of tacit collusion in duopolies than in triopolies or quadropolies.[2]

Similar results were obtained in Bertand oligopoly experiments. Fouraker and Siegel (1963), Dufwenberg and Gneezy (2000), Orzen (2008), Davis (2009) and Fonseca and Normann (2012) found less competitive behavior and higher profits in Bertrand duopolies, compared to triopolies or quadropolies. A meta-analysis and additional experiments in

[2]A slightly different design was used by Friedman et al. (2015), who compared Cournot duopolies to triopolies with a unit elastic demand function in a low information environment and 1200 4-second rounds. By the end of the game, output converged to the collusive outcome in the duopoly, but not in the triopoly. Oechssler et al. (2016) replicated the study using a standard linear demand function and found more collusive behavior in duopolies than in quadropolies.

Horstmann et al. (2018) corroborate these results.

The origins of the group size effect have been studied in other games. Isaac and Walker (1988) noted that in public goods games, an increase in group size would decrease the marginal per capita return of contributions to a public good. A change in the group size could therefore be divided into a "pure" group size effect and the part of the effect driven by the changes in incentives to contribute. Isaac and Walker (1988) observed a significant decrease in contributions when the group size was increased from 4 to 10. However, the effect disappeared when the payoff function was corrected to keep MPCR constant across group sizes.[3] Similarly, we find a decrease in the group size effect in Cournot oligopoly when the payoff function is normalized to keep the incentives comparable.

2 Experimental Design

2.1 Payoff Function

We study a symmetric n-firm Cournot oligopoly. Each firm $i \in N$ simultaneously chooses output q_i. Price p_i is determined by a linear inverse demand function:

$$p_i(q_i, q_{-i}) = \max\left(0, 81 - \left(q_i + \theta \sum_{j \in (N \setminus i)} q_j\right)\right) \quad (1)$$

where q_{-i} denotes the outputs of firms other than i.

The sole difference compared to the standard Cournot oligopoly implementation (e.g. Bigoni and Fort, 2013, Huck et al., 2004) is the addition of θ, which could be interpreted as the degree of product differentiation (Vives, 1984, Horstmann et al., 2018). If $\theta = 1$, as is commonly assumed, products are homogeneous and the market price is common for all

[3]Follow-up studies found that the group size may have a significant effect, although it depends on the value of MPCR and the nature of the game; for example, see Isaac et al. (1994), Barcelo and Capraro (2015), Nosenzo et al. (2015), Zelmer (2003).

firms. Otherwise, prices are different for each firm.

We assume that the marginal cost of production is equal to one, so that the cost function is $C(q_i) = q_i$. Then profits obtained by i are

$$\pi_i(q_i, q_{-i}) = (p_i(q_i, q_{-i})q_i - q_i)s - FC$$

The s parameter scales the payoffs to equalize equilibrium payoffs across games with a different number of firms. In previous research, a change in equilibrium payoffs due to a different group size was corrected using different exchange rates (e.g. Huck et al., 2004, Bosch-Domènech and Vriend, 2003). Instead, we use an explicit scaling parameter to increase transparency and maintain the same order of magnitude of payoffs across treatments, preventing any treatment effects due to the salience of payoffs. We set the fixed cost $FC = -130$, providing a subsidy that prevents negative payoffs and generates positive payoffs in the Walrasian equilibrium.[4]

Three key outcomes are typically studied in Cournot oligopoly games: collusive outcome, at which the sum of payoffs earned by all players is maximized; Nash equilibrium, at which all players best-respond to the action profile of all other players and Walrasian equilibrium, at which all players maximize their relative profits. We calculate the symmetric outcomes of interest using the standard procedure. In a symmetric Nash equilibrium, $q_i^N = \frac{80}{2+\theta(n-1)}$. In a symmetric collusive outcome, $q_i^C = \frac{80}{2+2\theta(n-1)}$. In a symmetric Walrasian equilibrium,[5] $q_i^W = \frac{80}{1+\theta(n-1)}$. There sometimes are asymmetric collusive outcomes and asymmetric Nash equilibria; in all asymmetric Nash equilibria, the total output is

[4] The choice of the scaling parameters and the strategy space also ensure that in the normalized treatments all the payoffs are three-digit numbers (i.e. within the range 100-999), therefore no part of the payoff space is particularly salient.

[5] With the parameters used in the experiment, the Walrasian equilibrium is at the point where price equals marginal cost. In general, the symmetric relative payoff maximization point could be different for small values of θ, i.e. if $\theta(n-1) < 1$.

equal to the total output produced in the symmetric equilibrium.[6]

2.2 Treatments

We aim to understand why output is more collusive in markets with fewer firms. Based on the findings from previous experiments and the insights from competition policy, we identify three potential explanations for this effect:

1. **In smaller markets, it is easier to coordinate.** In competition policy, market concentration is controlled in part due to a belief that "the presence of many competitors tends to make it more difficult to sustain coordination (...)".[7] This belief is supported by the results from economic experiments. Smaller groups manage to coordinate on higher prices and receive higher profits in Bertrand oligopoly (Dufwenberg and Gneezy, 2000; Davis, 2009; Fonseca and Normann, 2012), coordinate on the efficient equilibrium in a minimum effort game (Van Huyck et al., 1990) and sustain cooperation in voluntary contribution mechanism games (Nosenzo et al., 2015). The success of smaller groups could be explained by the use of "language of coordination" (Davis, 2009), which allows participants to signal the intentions to collude and identify deviations from collusive agreements (Masiliūnas, 2017). We will call this explanation the "pure number effect" (following Isaac and Walker, 1988), as it is driven purely by the group size, rather than by the change in incentives or other elements of the game.

[6] In the normalized treatments with the parameters used in the experiments and a discrete strategy space, there are 4 asymmetric equilibria in markets with two firms: (7,9), (9,7), (0,16), (16,0). There are 18 asymmetric equilibria in markets with three firms and 18 in markets with four firms. The average output is the same in all equilibria. Multiplicity of equilibria is common in Cournot oligopoly: for example, in the standard treatments, there are 2 asymmetric equilibria in two-firm markets, 6 equilibria in three-firm markets and 18 equilibria in four-firm markets.

[7] U.S. Department of Justice & Federal Trade Commission, Commentary on the Horizontal Merger Guidelines (2006), available at http://www.justice.gov/atr/public/guidelines/215247.pdf. Also see Ivaldi et al. (2007).

2. **In smaller markets, there are more incentives to collude or less incentives to deviate from collusive agreements.** In competition policy, mergers are expected to increase the individual incentives to increase prices, commonly measured using the value of diverted sales or an index of upward pricing pressure (Shapiro, 2010, Farrell and Shapiro, 2010). This "unilateral effect" is typically explained by the merged firm recapturing some of the sales lost due to a price increase by selling more substitute products (Jaffe and Weyl, 2013). The stability of collusive agreements in infinitely repeated Cournot experiments depends on the incentives to make a collusive agreement and the incentives to deviate from it. In standard Cournot oligopoly, firms have more incentives to collude, but also more incentives to deviate from the collusive agreement when there are more firms in the market. Overall, larger markets require a higher minimum discount factor to sustain collusion in an infinitely repeated game (Friedman, 1971). Our experiments follow the previous literature and use finitely repeated games, therefore collusive equilibria do not exist; however, we will show that changes in incentives predict more collusive output in smaller markets.

3. **In smaller markets, more choices are classified as collusive because the collusive outcome is closer to the focal options.** Preferences and beliefs depend on the choice set and item's location in the set; for example, there is a preference for options in the middle (Valenzuela and Raghubir, 2009; Chang and Liu, 2008) or for the multiples of 10 or 100. In standard Cournot oligopoly, the focality of the key outcomes, their location on the strategy space and the distance between them change as a result of a different market size. These changes could affect the behavior of boundedly rational participants and therefore more choices might be misclassified as collusive in smaller markets.

We designed the "normalized" treatments to keep the incentives and the focality of actions the same across markets of different size to identify the pure number effect. We also ran experiments with the "standard" design to compare the number effect between the two designs using the participants from the same subject pool and identical experimental procedures.

Table 1 summarizes the key differences between the five treatments. The difference between standard and normalized treatments lies in how the output of the opponents is aggregated, which is determined by the θ parameter. In standard treatments, we set $\theta = 1$, regardless of the market size, as is common in the previous literature (e.g. Huck et al., 2004, Roux and Thöni, 2015, Horstmann et al., 2018, Oechssler et al., 2016). As a result, a change in the market size does not affect the profit of firm i if the output of firm i and the *sum* of output of all other firms are held constant.[8] In normalized treatments, we set $\theta = \frac{1}{n-1}$, therefore market size does not affect the profits of i if the output of i and the *average* output of all other firms are held constant. This small difference in the aggregation of opponents' output has important consequences on how the incentive structure responds to changes in market size.

In the standard treatments, market size affects the output and payoffs in the three key outcomes: Nash equilibrium, collusive outcome and Walrasian equilibrium. First two columns of Table 1 summarize these differences in markets with 2 and 4 firms (S2 and S4). Nash equilibrium payoffs are held constant using a scaling parameter ($s = 1$ in 4-firm markets, $s = 0.36$ in a 2-firm markets), but the payoffs in the collusive outcome and the most profitable deviation from it are different. A common measure of the incentives to collude is the Friedman index (Friedman, 1971), defined as $F = \frac{\pi(C)-\pi(N)}{\pi(D)-\pi(C)}$, where $\pi(D)$ is the payoff in the most profitable unilateral deviation from the collusive outcome. Table 1 shows that in the standard treatments, the incentives to collude decrease in the group size.[9]

[8]Note than in the experiment, we set s to different values in 2 and 4 player markets, in line with the previous literature, therefore the profits in one treatment will be multiplied by a scalar; however, multiplication does not affect the best response.

[9]Friedman index values in our experiment are identical to the values in standard experiments that

Table 1: Parameter values and key outcomes in each treatment.

Treatment	S2	S4	N2	N3	N4
n	2	4	2	3	4
θ	1	1	3	1.5	1
s	0.36	1	1	1	1
Strategy space	0-50	0-50	8-24	8-24	8-24
q_i^C	20	10	10	10	10
q_i^N	26.67	16	16	16	16
q_i^W	40	20	20	20	20
$\pi(C)$	418	530	530	530	530
$\pi(N)$	386	386	386	386	386
$\pi(W)$	130	130	130	130	130
Friedman index	0.89	0.64	0.64	0.64	0.64
# participants	36	60	96	72	96
# markets	18	15	48	24	24

In the normalized treatments, the locations of the three key outcomes and the payoffs are invariant to the market size.

The second difference between standard and normalized treatments is in the strategy space. In standard, firms could choose output from 0 to 50. In normalized, the strategy space was restricted to 8-24, for several reasons. First, we aimed to increase the accuracy of categorization by having the three key outcomes evenly spaced across the strategy space. If all three outcomes were located at the bottom of the strategy space (e.g. as in S4), there is more scope for exceeding the equilibrium prediction and the average output might appear to indicate competitiveness. Instead, in normalized treatments the Nash equilibrium was placed in the middle of the strategy space and the other two outcomes were located at a similar distance from it. However, we did not want any outcomes to be located at either end of the strategy space, so as not to misclassify extreme choices as either perfectly collusive or competitive. It was also necessary to bound the strategy space from below to decrease the attractiveness of collusion by an alternating play of asymmetric outcomes.[10]

manipulated group size in Cournot oligopoly, e.g. Huck et al. (2004).

[10]In N2, the symmetric collusive outcome is (10,10), generating a payoff of 530 ECU per firm. However, total payoffs would be maximized in the output profile (8, 16), with an average payoff of 546 EU per firm.

Finally, the strategy space in normalized treatments was labeled such that the key outcomes would not be focal. For example, in S4 treatment, which uses the same parameters as the previous literature, the collusive and competitive outcomes are at focal locations (respectively 10 and 20), and it is found that these outcomes are commonly chosen (Bigoni and Fort, 2013). It is unclear whether these values are chosen because they were focal, or because participants converge to the collusive or competitive outcome. To simplify the explanation of the game and separate the key outcomes from focal values, we re-labeled the strategy space, using strategy space from 0 to 16 to represent output 8-24. The mapping was performed in two ways: in "increasing" treatments, higher output was represented by higher numbers in the strategy space (i.e. output of 8 was labeled as "0" and 24 was labeled as "16"). In "decreasing" treatments, higher output was represented by lower numbers (i.e. output of 8 was labeled as "16" and 24 was labeled as "0"), reversing the strategy space. This labeling ensures that the key outcomes are never at focal locations (in increasing treatments, key outcomes are at 2, 8 and 12; in decreasing treatments they are at 4, 8 and 14). Running some treatments with a reversed strategy space helps to further evaluate the importance of focality and to eliminate it's effect when classifying choices (for more details, see Appendix A).[11] In the data analysis section, we map the choices made by participants back into the output of the original Cournot payoff function and pool increasing and decreasing treatments.

In N3, the asymmetric collusive outcome is (8, 8, 16), with a profit of 535.33 EU per firm. In N4, all output profiles in which total output equals 40 generate the same payoff of 530 ECU per person. If the strategy space was not bounded from below, the payoff difference between symmetric and asymmetric collusive outcomes would be much larger, and cause different patterns of behaviour in 2 and 4 firm markets. For example, the asymmetric collusive outcome in an unbounded 2-firm treatment is (4, 24), with a payoff of 810 ECU per person. The lower limit of 8 makes the asymmetric collusive outcome less attractive, while keeping the symmetric collusive outcome in the interior of the strategy space. To completely eliminate the asymmetric collusive outcome, the lowest available output would have to be 10. In the data, we do not observe any successful attempts of asymmetric collusion.

[11] We find evidence that focal outcomes are more commonly chosen. For example, action labeled as "10" is the most commonly chosen action in N3I (increasing treatment with a market size of 3), second most common in N4I and N3D, third most common in N2I, N2D and N4D. When the same action is labeled as "6", it is chosen less than half of the times. If we hadn't run the treatments with a reversed strategy space, the frequency of competitive outcomes would likely be overestimated.

2.3 Other Design Details

The stage game was repeated 20 times under partner matching. An alternative random matching protocol would have made it very difficult to study collusion. Participants had 30 seconds to make their decision in each round (as in Bigoni and Fort, 2013). If no decision was made within the time limit, the output chosen in the previous round was implemented. If no decision was made in round 1, output was drawn at random from a uniform distribution.[12]

Original instructions were in French; Appendix D provides a complete English translation. Instructions were identical in all treatments and framed using neutral language.[13] Participants could learn about the incentive structure using either a payoff table or a payoff calculator (examples of the information seen by the participants are displayed in figures F.1 and F.2 in Appendix F). The payoff table listed participant's payoffs for some combinations of chosen output and the average output of the opponents: 441 combinations in standard treatments (21x21 action profiles, 0-50 in increments of 2.5) and 289 combinations in normalized treatments (17x17 action profiles, 0-16 in increments of 1). The payoff calculator could be used to compute the payoff for any combination of own output and the average output of the opponents.

In each round other than the first one, participants had access to two additional tools. The "output-payoff graph" visually displayed the previous round output-payoff combinations of all the firms in the market (see Friedman et al., 2015 for a similar design). This information is needed to imitate the best, which could lead to convergence to the Walrasian equilibrium (Vega-Redondo et al., 1997). The second tool was a table that listed the history of chosen output, payoff and average output of the opponents in the previous rounds.

[12] The decision was not made within the time limit 3.5% of the time, primarily in the first two rounds. In the analysis part, we use all the decisions, although excluding the decisions that were not made explicitly does not change the overall results.

[13] For a discussion about how the use of neutral language rather than the more commonly used economic framing affects preferences and beliefs in Cournot oligopoly, see Masiliūnas and Nax (2020).

This information would be used to make decisions from experience (e.g. reinforcement or belief learning). Participants could switch between the four tools at any time, and we tracked how much time was spent using each tool, just as in Bigoni and Fort (2013). This process data provides additional insight into how the decision making process is affected by the market size.

After 20 rounds, participants continued the experiment with different games for another 20 or 40 rounds (depending on the treatment). This data was collected to study learning transfer and is used in a separate paper. Table C.1 in Appendix C shows the structure of the entire experiment. In this paper, we use only choices from the first 20 rounds. Participants were aware that the experiment will contain multiple parts, but did not know how many parts there will be and what type of games will be played. In each part, participants were matched with different opponents, therefore they never interacted with their opponents from the first part again. One round from each part was randomly selected at the end of the experiment, and the earnings from these rounds were added up and paid to the participants in cash.[14]

Additional information was collected at the end of the experiment. We elicited social preferences using the Social Value Orientation slider measure (Murphy et al., 2011, using the z-Tree implementation by Crosetto et al., 2019). We also measured the cognitive abilities using a part of the advanced version of Raven's Progressive Matrices task (Raven and Court, 1998). In this test, participants had 10 minutes to solve 16 tasks. After these tasks, we collected the age, gender and year of study of the participants.

In total, 360 participants took part in experiments. The numbers of participants and markets in each treatment are shown in Table 1. In normalized treatments, exactly one half of the participants in each treatment took part in "increasing" treatments and the other half in "decreasing" treatments. We collected more observations for the normalized

[14]Earnings were denominated in ECU and exchanged to cash using rate 150 ECU = 1 euro.

treatments because the standard treatments have already been studied in the previous literature.

All experiments were run in the LEEN laboratory of the Université Côte d'Azur in May and October 2018. The experiments took on average 75 minutes and participants on average received 14.2 euros. Participants were recruited using ORSEE (Greiner, 2015) and experiments were programmed using z-Tree (Fischbacher, 2007).

3 Quantal Response Equilibrium Predictions

We have shown that the market size interacts with the incentive scheme differently in standard and in normalized treatments. This section quantifies how these differences in incentives are predicted to affect the collusion rates under bounded rationality, as modeled by the Quantal Response Equilibrium (McKelvey and Palfrey, 1995).

QRE requires consistency between actions and beliefs, but the responses to beliefs are noisy, therefore all actions have a positive probability to be played. In a game with n players, a set of players N and a set of pure strategies $Q_i = \{q_1, \ldots, q_m\}$, denote the set of all probability measures on Q_i by $\Delta_i = \{p_1, \ldots, p_m\}$ and the set of all probability measures on $\times_i Q_i$ by $\Delta = \{\Delta_1, \ldots, \Delta_n\}$. We will use shorthand notation $p = (p_i, p_{-i})$ for any $p \in \Delta$, where p_i is the mixed strategy of player i and p_{-i} is the mixed strategy profile of all other players. The probability with which player i chooses action q_k is $p_i(q_k)$. The expected payoff that i obtains by choosing q_k is denoted by $\pi_i(q_k, p_{-i})$. Nash equilibrium assumes that each player chooses the action with the highest expected payoff. Instead, QRE assumes that participants are maximizing their decision utility $u_i(q_k, p_{-i})$, equal to the sum of the expected payoff and the noise term:

$$u_i(q_k, p_{-i}) = \pi(q_k, p_{-i}) + \varepsilon_{ik} \tag{2}$$

If each of the stochastic terms ε_{ik} is independently drawn from a type-I extreme value distribution with parameter λ (McFadden, 1981) and each player chooses the action that generates the highest decision utility, the probability that i will play q_k can be calculated using a noisy best-response function $\sigma_i(q_k, p_{-i})$, defined as:

$$\sigma_i(q_k, p_{-i}) = \frac{e^{\lambda \pi_i(q_k, p_{-i})}}{\sum_{q_j \in Q_i} e^{\lambda \pi_i(q_j, p_{-i})}} \tag{3}$$

The logit QRE is a probability distribution $p \in \Delta$ that satisfies $p_i(q_k) = \sigma_i(q_k, p_{-i})$, for all $i \in N$ and $q_k \in Q_i$ (McKelvey and Palfrey, 1995). In other words, QRE requires the mixed strategy of each player to be a noisy best-response to the mixed strategy profile used by all other players.

Parameter λ measures precision, or sensitivity to expected payoff differences. If $\lambda = 0$, all actions are chosen with equal probabilities. A positive λ indicates that actions that generate higher expected payoffs are chosen more often. If $\lambda \to \infty$, there is no error and players always choose the action with the highest expected payoff, therefore QRE reduces to the Nash equilibrium.

Since the closed-form expressions of logit QRE are generally unknown, we calculate QRE using the tracing procedure from Turocy (2005), implemented using Gambit software (McKelvey et al., 2015). The calculations are performed using a discretized strategy space of 51 strategies (0, 1, ..., 50) in standard treatments and 17 strategies (8, 9, ..., 24) in normalized treatments. Payoffs used in the calculations are converted into monetary euro amounts (in experiments, the exchange rate was 150 ECU = 1 euro).

3.1 QRE in Standard Treatments

We start by evaluating how QRE predictions respond to changes in market size in standard treatments and move to the normalized treatments in subsection 3.2. First, we illustrate

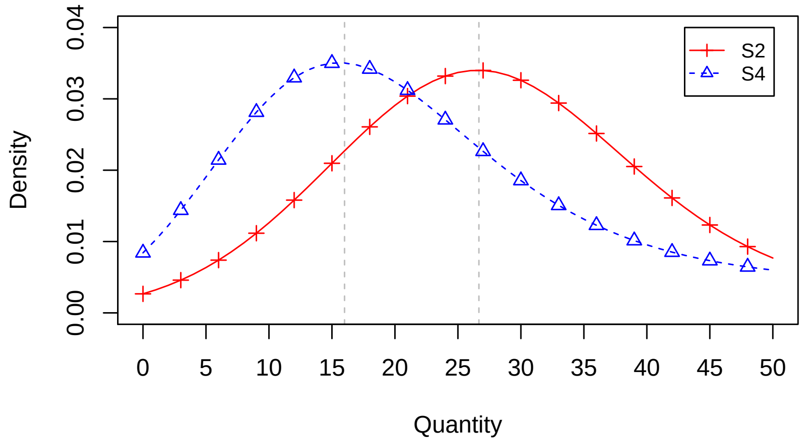


Figure 1: QRE distribution with $\lambda = 1.5$ in standard treatments. Vertical dashed lines indicate Nash equilibrium in S4 (16) and S2 (26.7).

the QRE predictions by calculating the choice probabilities for a specific value of the noise parameter ($\lambda = 1.5$, very close to the value in the best-fitting QRE model estimated in section 5.1) and then show how the predictions respond to changes in λ.

Figure 1 shows the calculated QRE probability distributions in treatments S2 and S4. Dashed lines mark the corresponding Nash equilibrium output (26.7 and 16). Both distributions are centered around the Nash equilibrium, but more choices exceed the Nash equilibrium prediction in S4 than in S2 (59%, compared to 51%), largely because in S4 the Nash equilibrium is located at the bottom of the strategy space. Consequently, the average QRE output is very close to the Nash equilibrium in S2 (26.8, compared to 26.7), but exceeds the Nash equilibrium in S4 (20.6, compared to 16).

Next, we introduce two measures to quantify the degree of collusion and compare them across a range of parameter values. The first measure is the ratio of average output to Nash equilibrium output (Huck et al., 2004), calculated as $r = \frac{\hat{q}(\lambda)}{q^N}$, where $\hat{q}(\lambda)$ is the average output in a QRE with parameter value λ, and q^N is the Nash equilibrium prediction. Figure 2 plots r in S2 and S4 for values $\lambda \in [0, 10]$. If the sensitivity to payoff differences is high, QRE approaches Nash equilibrium and thus $r \to 1$. If the sensitivity is low, average

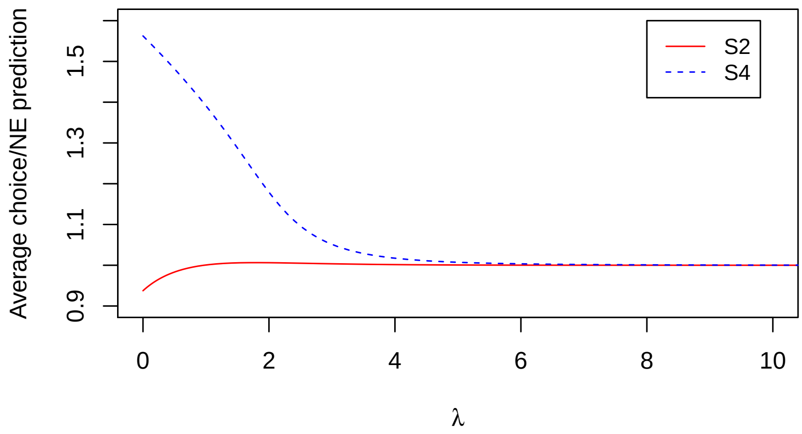


Figure 2: Ratio of average QRE output to Nash equilibrium output in standard treatments with 2 and 4 firm markets.

output is below Nash equilibrium in two-firm markets and above it in four-firm markets, predicting more collusive behavior in smaller markets.

The second measure is the frequency of collusive or competitive choices. It is common to classify choices by identifying the outcome that is closest to the chosen output (see Huck et al., 2004). We partition the strategy space into actions that are closest to either collusive outcome, Nash equilibrium or Walrasian equilibrium, and compare the QRE predictions about the frequency of each category across treatments. We find that smaller markets are predicted to have more collusive and less competitive choices, for a wide range of λ values (Figure 3).

Overall, QRE predicts more collusive behaviour in smaller markets when the market size is manipulated using the standard Cournot payoff function. This finding indicates that the experimental result of lower collusion rates in smaller markets could in principle be driven by changes in the incentive structure or the locations of the key outcomes.

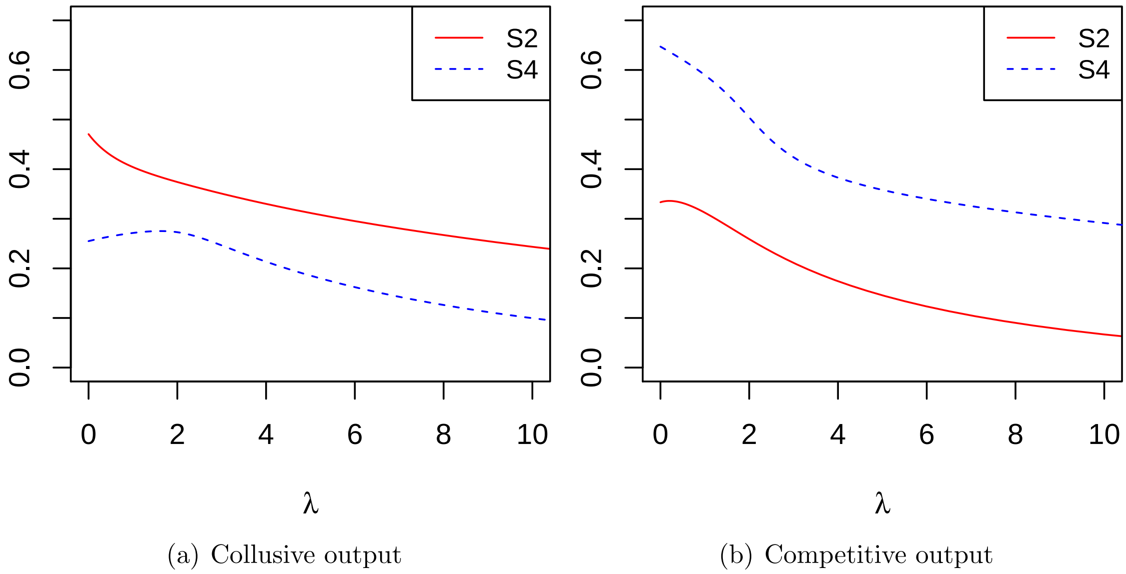


Figure 3: Fraction of choices classified as collusive (closer to the collusive outcome than to the other two outcomes) or competitive (closer to the Walrasian equilibrium than to the other two outcomes) in standard treatments.

3.2 QRE in Normalized Treatments

Next, we calculate QRE in normalized treatments. The normalized payoff function was designed to keep the incentives similar across different market sizes. We therefore expect smaller differences in predicted collusion rates between normalized treatments.

Figure 4 illustrates the QRE distribution for $\lambda = 1.5$. The markets with 3 and 4 firms are similar, but the 2-firm market is somewhat shifted towards more competitive output. This shift is caused by a difference in how the distribution of beliefs about average output is constructed. In N2, beliefs coincide with the QRE distribution, since players face only one opponent. In N3 and N4, the distribution of beliefs about the average output will have a lower variance because extreme values of average output are less likely in larger markets. Higher variance of the belief distribution in N2 makes higher output more attractive because such output is on average more profitable when the average output chosen by the opponents is more extreme (see the payoffs in Table E.1, Appendix E).

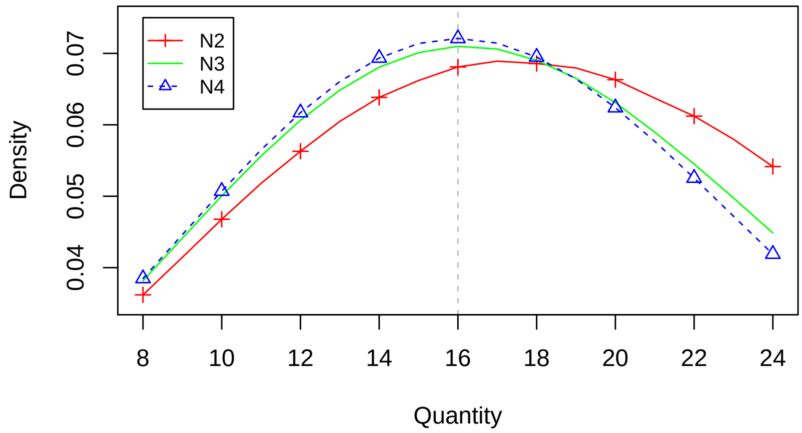


Figure 4: QRE distribution in normalized treatments at $\lambda = 1.5$. Vertical dashed line indicates Nash equilibrium output.

We quantify collusion using the two measures introduced in the previous section: a ratio of average output to Nash equilibrium output and the frequency of output classified as either collusive or competitive. Figures 5 and 6 show how these measures are predicted to depend on the market size.

Figure 5 plots the ratio of average QRE output to Nash equilibrium output. In contrast to the standard treatments (Figure 2), we find that smaller markets are predicted to be slightly less collusive. However, the treatment difference is much smaller compared to the standard treatments: the ratio differs by at most 5%, in contrast to the differences of up to 60% in the standard treatments.

Figure 6 shows that the slightly less collusive output in markets with two firms is primarily driven by a higher frequency of competitive choices in this treatment. The predicted fraction of collusive choices is nearly identical across the different market sizes, and much smaller than the differences in the standard treatments.

Overall, the QRE prediction that smaller markets will be more collusive holds only in the standard but not in the normalized treatments. The comparison of the group size effect under both designs can therefore identify whether the results found in the Cournot

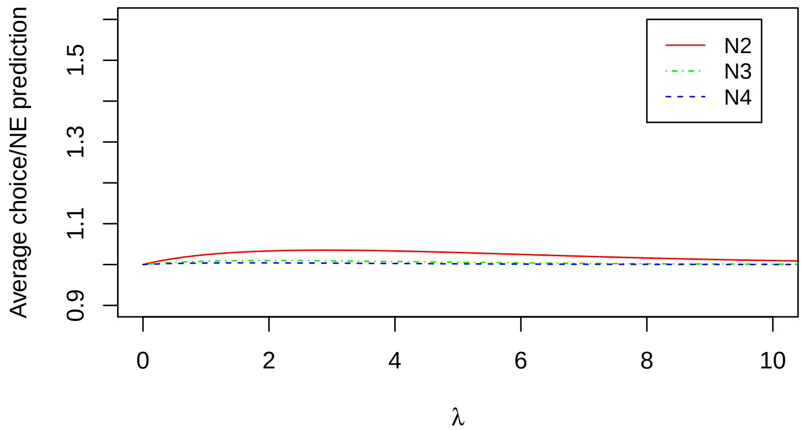


Figure 5: Ratio of average QRE output to Nash equilibrium output in normalized treatments with 2, 3 and 4 firm markets.

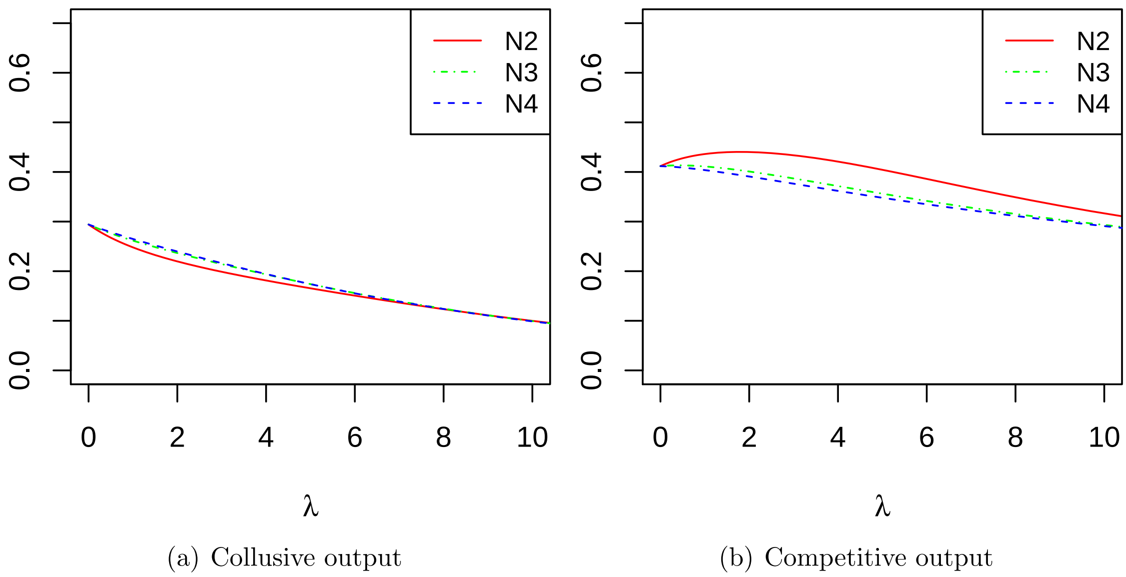


Figure 6: Fraction of choices classified as collusive (closer to the collusive outcome than to the other two outcomes) or competitive (closer to the Walrasian equilibrium than to the other two outcomes) in normalized treatments.

literature are primarily driven by the pure number effect, or by the changes in the incentive structure or the focality of key outcomes.

4 Results

First, we will compare the two designs (standard and normalized) in whether they alter the effect of the market size on the aggregate levels of collusion. Afterwards, we will compare the distributions of output and classify behavior to identify how the market size affects the frequency of collusive output.

4.1 Aggregate Output

In the standard treatments, theoretical predictions change with the market size, therefore average output ($\bar{q}_i$) needs to be normalized before treatments can be compared. We do so using two measures of collusion: the ratio of actual to predicted output and a collusion index.

Ratio of actual to predicted output. A simple way to normalize output is to calculate the ratio of chosen output to Nash equilibrium output: $r = \bar{q}_i/q_i^N$. Values below 1 indicate that output is more collusive than the equilibrium prediction, values above 1 indicate competitive output. Figure 7 shows the dynamics of the across markets average r. To compare across treatments, we calculate r separately for each market, aggregated either across all rounds (1-20) and only the last five rounds (15-20), following the convention in the literature (e.g. Huck et al., 2004). We use a Mann-Whitney U (MWU) test to test whether the ratio of output is significantly affected by market size.

In standard treatments, there is a significant difference between the markets with 2 and 4 firms (MWU $p < 0.0001$ if all rounds are used, $p = 0.001$ if last 5 rounds are used). This result replicates the finding from the literature that smaller markets tend to be more

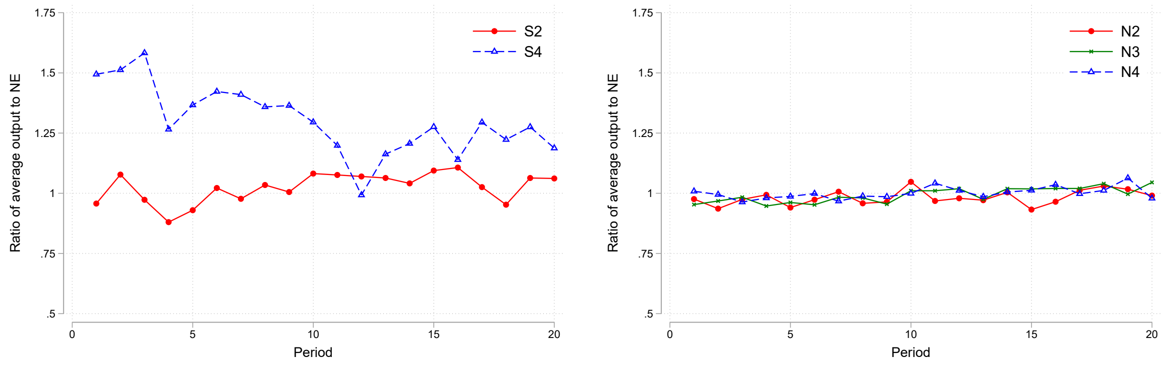


Figure 7: Ratio of average output to equilibrium output by treatment over time.

collusive (Huck et al., 2004). In the normalized treatments, there is no difference between markets with 2, 3 or 4 firms (MWU p-values for the pairwise comparison are at least 0.5826 if all rounds are used and 0.3123 if the last 5 rounds are used).

Collusion index. Used in Horstmann et al. (2018), Engel (2007) and Suetens and Potters (2007), the collusion index measures where the market output falls in the range between the collusive outcome and the Nash equilibrium or the Walrasian equilibrium: $\varphi^N = (\bar{q}_i - q_i^N)/(q_i^C - q_i^N)$ and $\varphi^W = (\bar{q}_i - q_i^W)/(q_i^C - q_i^W)$. Both indexes would be equal to 1 if market output was equal to the collusive output; the first index would be equal to 0 if output was equal to the Nash equilibrium, the second would be equal to 0 if output was equal to the Walrasian equilibrium. Note that in the normalized treatments, the treatment comparison is identical for all three indexes because the key outcomes are invariant to market size. We find that in the standard treatments, 2-firm markets are more collusive than 4-firm markets, as measured by φ^N (MWU $p = 0.0001$ for all rounds, $p = 0.0103$ for last 5 rounds) or by φ^W (MWU $p < 0.0001$ for all or only the last 5 rounds). In the normalized treatments, there is no difference between the markets with 2, 3 or 4 firms (MWU $p > 0.5786$ for all rounds and $p > 0.3123$ for the last 5 rounds).

From a policy perspective, it is important to know whether firms in more concentrated markets receive higher profits. Nash equilibrium predicts identical profits in all

Table 2: Three measures of collusion and profits across treatments. First number is the average in rounds 1-20, the number in brackets is the average in rounds 15-20.

Index	S2	S4	N2	N3	N4
r	1.02	1.30	0.98	0.99	1.00
	[1.05]	[1.23]	[0.99]	[1.02]	[1.02]
φ^N	-0.10	-0.80	0.05	0.02	-0.002
	[-0.20]	[-0.62]	[0.02]	[-0.06]	[-0.04]
φ^W	0.63	-0.08	0.43	0.41	0.40
	[0.60]	[0.03]	[0.41]	[0.36]	[0.37]
π_i	340.0	230.1	387.3	371.1	372.4
	[338.8]	[243.9]	[381.9]	[348.0]	[361.9]

five treatments. Table 2 shows that in the standard treatments, profits are significantly higher in 2-firm markets, both overall (MWU $p < 0.0001$) and in the last 5 rounds (MWU $p = 0.0001$). There is, on the other hand, no significant difference among three market sizes in the normalized treatments (the lowest MWU p-value is 0.2323 in all rounds and 0.1450 in the last 5 rounds).

Additionally, we evaluate the treatment effects using a panel data GLS regression with a random effect on the market level, taking into account the inter-temporal dependence of decisions as well as the dependence among the outputs of firms in the same market. Standard errors are clustered on the market level. As dependent variables, we use the three collusion indexes (r, φ^N, φ^W). The main independent variables are dummy variables indicating market size. For the normalized treatments, we also include an indicator of the strategy space labeling. A variable equal to the inverse of a round is included to capture changes in collusion due to experience. Table 3 shows that, in the standard treatments, markets with four firms are significantly less collusive than markets with two firms, for all three indexes. The coefficients of the inverse round variable indicate increasing collusion over time. In the normalized treatments, the 3-firm and 4-firm markets are not different from the 2-firm markets. The "decreasing labels" variable indicates treatments in which higher output was labeled with lower numbers. The coefficient of this variable is signifi-

Table 3: Random effects GLS regression. Standard errors clustered on the market level.

	Standard			Normalized		
	DV: r	DV: φ^N	DV: φ^W	DV: r	DV: φ^N	DV: φ^W
3-firm market	–	–	–	0.0109	-0.0291	-0.0175
				(0.42)	(-0.42)	(-0.42)
4-firm market	0.277***	-0.705***	-0.716***	0.0189	-0.0504	-0.0303
	(6.37)	(-4.96)	(-11.02)	(1.04)	(-1.04)	(-1.04)
1/Round	0.206**	-0.501*	-0.340***	-0.0382*	0.102*	0.0611*
	(2.49)	(-1.96)	(-2.68)	(-1.75)	(1.75)	(1.75)
Decreasing labels				-0.0903***	0.241***	0.144***
				(-5.19)	(5.19)	(5.19)
Constant	0.988***	-0.00840	0.695***	1.034***	-0.0907*	0.346***
	(32.33)	(-0.07)	(16.45)	(58.88)	(-1.94)	(12.30)
N	1920	1920	1920	5280	5280	5280

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

cantly different from 0, indicating the importance of labeling. The effect is largely driven by action "10" being focal; in the increasing treatments, this action is more competitive than the Nash equilibrium but in decreasing treatments it is more collusive. We note that the labeling of the strategy space has a significant and consistent effect on collusion rates, which is even stronger than the effect of market size.[15]

These results are robust to different specifications. Table C.3 in Appendix C shows the estimated treatment effects using data only from the last 5 rounds, when participants would have accumulated experience. The treatment effects remain qualitatively the same. Results also do not change if we remove decisions that were implemented because participants fail to make a decision within the time limit. In the normalized treatments, we ran regressions separately for the decreasing and increasing strategy space and found no significant number effects (Table C.2 in Appendix C). Including age and gender does not change the results, and coefficients for these two variables are not significantly different from zero.

Result 1. *In standard treatments, aggregate output is more collusive in smaller markets.*

[15] Further details on the effect of strategy space labeling are in Appendix A.

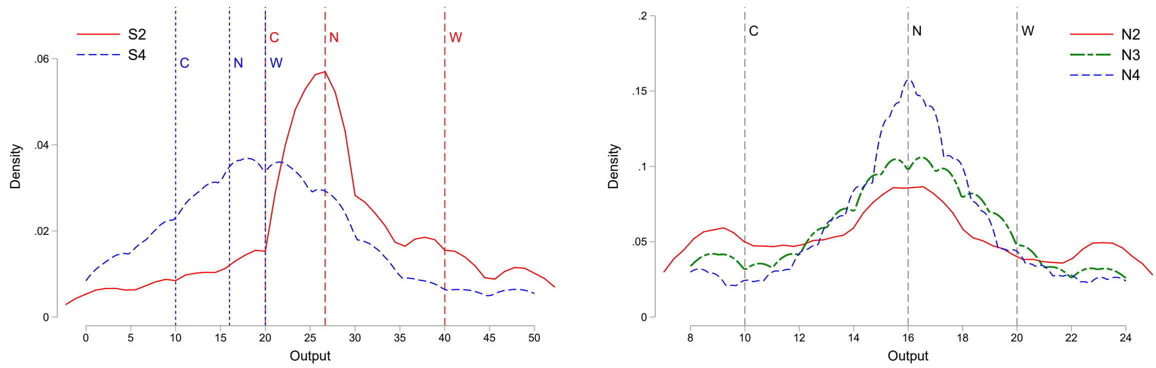


Figure 8: Kernel density of individual output, pooled across rounds.

In normalized treatments, market size has no effect on aggregate output.

4.2 Distribution of Output

Comparing treatments in terms of aggregate outcomes is important to compare our results to the previous literature and for policy reasons, as regulators care about the average profits and prices in the markets. But we are also interested in how and why the market size affects the distribution of output and the frequency of collusive output.

Figure 8 shows the kernel density estimate of individually chosen output in the standard and normalized treatments. In standard treatments, a larger market size decreases output but the adjustment is not as strong as predicted by Nash equilibrium. Consequently, the average output is close to the equilibrium prediction in 2-firm markets, but exceeds it in 4-firm markets. In normalized treatments, distributions are centered around the Nash equilibrium but the variance is notably lower in larger markets: the standard deviation of chosen output is 4.9 in N2, 4.2 in N3 and 3.7 in N4. We evaluate the statistical significance of these differences by calculating the standard deviation for each market and comparing their distribution across treatments. In normalized treatments, we find that the standard deviation is significantly higher in two-firm markets, compared to the markets of larger size (Mann-Whitney U test $p = 0.0041$ comparing N2 vs N3 and $p = 0.0017$ comparing

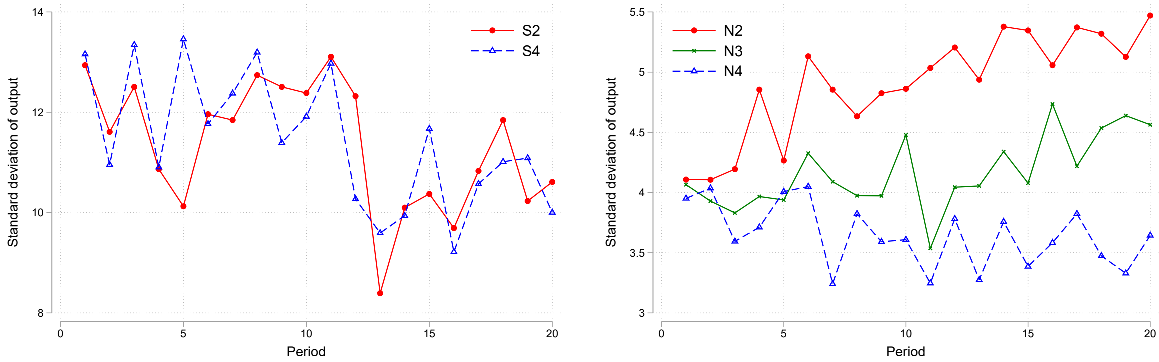


Figure 9: Standard deviation of choices by treatment over time.

N2 vs N4). In standard treatments, market size has no significant effect on the standard deviation. Figure 9 plots the evolution if the standard deviation, calculated using all choices in each round. In standard treatments, there is no difference between the two market sizes. In normalized, there is no difference at the start of the game, but a gap between the three treatments appears and grows over time.

Differences in the choice distributions affect the fraction of choices classified as collusive. We use two methods to identify the frequency of collusion. First, we use collusion counts (Waichman et al., 2014), defined as the number of rounds in which quantities are in the collusive region, that is closer to the collusive outcome than to the Nash equilibrium. Unlike Waichman et al. (2014), we perform the classification using individual rather than total market output, because aggregating choices over a larger number of firms decreases the likelihood that the group would be classified as collusive.[16] The average number of rounds in which individual output is classified as collusive is 5.1 in S4 and 5.2 in S2, a difference that is not significant (MWU $p = 0.49$). In normalized treatments, collusion counts go up from 3.0 in N4 to 4.1 in N3 and 5.8 in N2; there is significantly more collusion in 2-firm

[16]Classification using the market output replicates the usual finding of more collusive choices in smaller markets: in standard treatments, the average collusion counts go up from 5.1 to 10.3; in normalized treatments, from 3.7 to 8.1 to 9.0. The difference between 2 and 4-firm markets is significant in both standard (MWU $p = 0.0084$) and normalized treatments (MWU $p = 0.0009$). The 3-firm market is not significantly different from the other two.

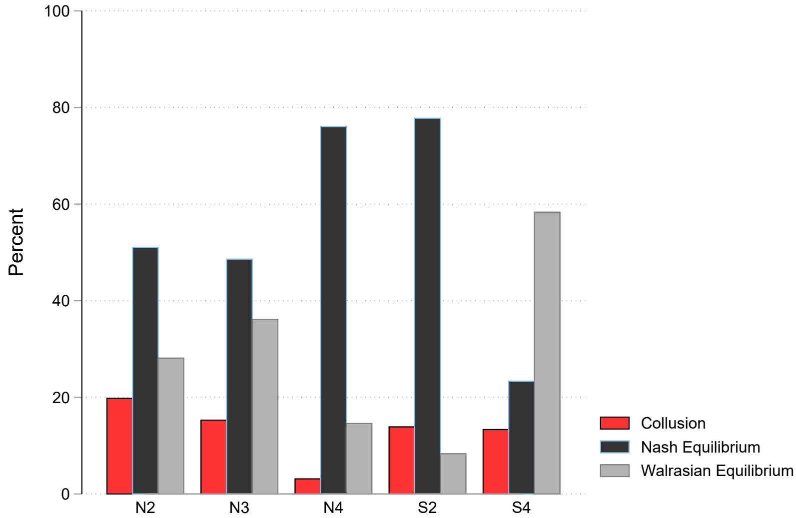


Figure 10: Classification of choices according to which outcome they are closest to. Data from last 5 rounds.

markets than in 3-firm markets (MWU $p = 0.0299$) or 4-firm markets (MWU $p = 0.0018$).

The comparison of collusion counts indicates a higher incidence of collusion in smaller markets, in contrast to the results based on aggregate output. The reason behind this difference can be seen by inspecting the output distributions. In standard treatments, a decrease in market size primarily lowers the frequency of above-equilibrium output, lowering the average output but not increasing the share of collusive choices. In normalized treatments, an increased variance in smaller markets does not change the average output but increases the frequency of choices that are close to the collusive output.

Next, we classify individual output based on the outcome it is closest to. We follow Huck et al. (2004), but use output from individual firms rather than the total market output. We aim to identify the outcome to which choices converge, therefore we use the average output of each firm in the last 5 rounds. Figure 10 shows that in standard treatments, a larger market size increases the fraction of choices classified as Walrasian (competitive) but has no effect on the frequency of collusive choices. In normalized treatments, a larger market size decreases the frequency of collusive choices but has a non-monotonic effect on the frequency of Walrasian or Nash choices.

Result 2. *In normalized treatments, the frequency of collusive output and the variance of output are higher in smaller markets. In standard treatments, market size has no effect on the frequency of collusion or the variance of output.*

5 Model Estimation

Our standard treatments replicate the previous literature, finding more collusion in smaller markets. In the normalized treatments, we find no effect of the market size on the aggregate output or profits. However, in these treatments smaller markets tend to have a higher variance of output, an effect that is not found in the standard treatments. This section will test whether models of bounded rationality can explain these data patterns.

5.1 Quantal Response Equilibrium

First, we test whether the data patterns can be explained by QRE. In section 3, we showed that QRE correctly predicts that the number of competitors makes output less collusive in the standard treatments, but does not affect it in the normalized treatments. But QRE also makes a prediction about the entire distribution of choices, therefore we can test whether it can explain the differences in variance observed in the normalized treatments. We estimate the logit QRE using a method adapted from Bajari and Hortacsu (2005): we calculate the noisy best-response to the empirical choice distribution, and use a grid search procedure to find a precision parameter (λ) that maximizes the likelihood of data in the experiment. This method produces an unbiased estimate of the precision parameter under the assumption that QRE is correct because in QRE beliefs would coincide with the empirical distribution. In practice, QRE will never fit the data perfectly, so there will be

Table 4: Goodness of fit and estimated parameter values in QRE.

	Separate estimation		Combined estimation	
Treatment	$\hat{\lambda}$	LL	$\hat{\lambda}$	LL
S2	1.395	-2739.97	1.497	-2740.34
S4	1.535	-4561.56	1.497	-4561.65
N2	-0.292	-5436.98	1.497	-5531.25
N3	2.507	-4014.05	1.497	-4024.05
N4	4.686	-5161.69	1.497	-5281.54
Total		-21914.25		-22138.83

a discrepancy between this method and the tracing procedure we used to compute QRE in section 3.[17]

When estimating QRE, we either fit a separate model for each treatment, or fit a single model, combining data from all five treatments. In the literature, it is common to estimate a separate model for each game (e.g. Lim et al., 2014), but it implicitly assumes that the sensitivity to payoff differences is changing with market size. Since the normalized treatments were designed to be comparable with each other, allowing the noise parameter to vary with the market size would lead to overfitting. If very different values of λ are needed to explain the number effect, results would be driven not by the elements in QRE, but by some other unmodeled factors. The second method therefore tests whether all the treatment differences could be explained by a model with a single value of the noise parameter.[18]

Table 4 provides an overview of the estimated parameter values and goodness of fit (log

[17]There are two main benefits of using the estimation procedure from Bajari and Hortacsu (2005). First, it gives the flexibility to perform a combined estimation using multiple treatments with different payoff function, which is not possible with the standard tracing procedure. Second, the computational complexity of the estimation procedure is greatly reduced, since it is no longer necessary to compute the fixed point for a large number of parameter values. Computational complexity is especially problematic for the games with many participants and a large strategy space, such as our S4 treatment (51 strategies for each of the 4 participants).

[18]We also estimated QRE separately for normalized and standard treatments, requiring λ to be the same across different market sizes but allowing them to differ between standard and normalized treatments. This is justified by the important differences between standard and normalized treatments, such as the range of payoffs or the size of the strategy space, which affect the QRE predictions. In practice, the estimates are very close to the combined estimation. We report these results and goodness of fit in Appendix C.

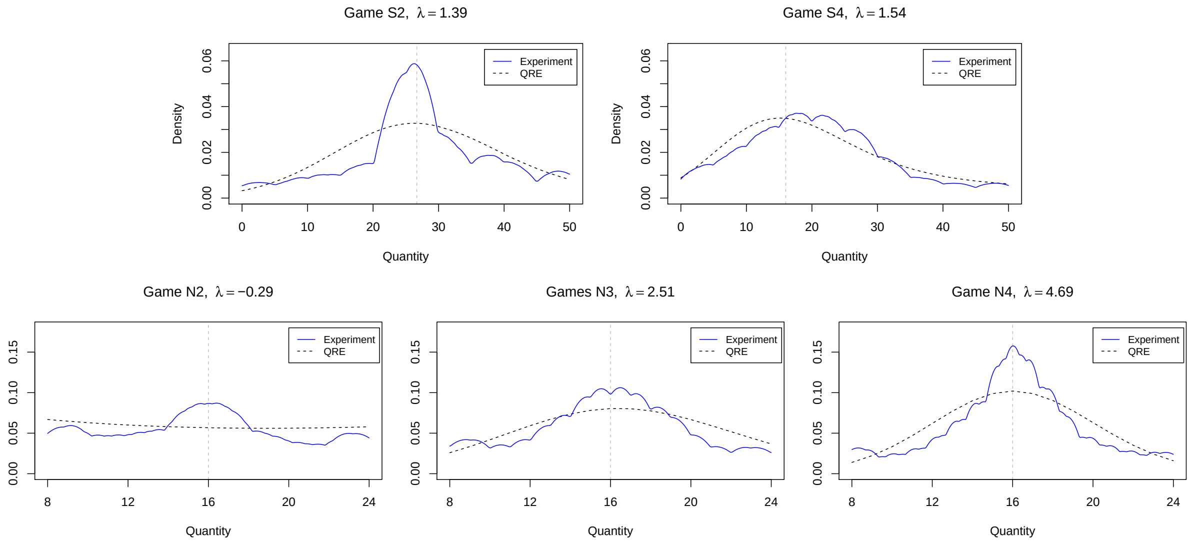


Figure 11: Best-fitting QRE estimated separately for each treatment (vs kernel density).

likelihood) for each method. When a separate model is fit for each game, the estimated values of the noise parameter are similar in the two standard treatments ($\hat{\lambda} = 1.4$ and $\hat{\lambda} = 1.5$), but very different in the three normalized treatments. That indicates that QRE can explain the number effect in standard treatments (result 1), but cannot explain the increased variance in smaller markets of the normalized treatments (result 2), unless it is assumed that sensitivity to expected payoff differences significantly decreases in smaller markets. In fact, the estimated value of λ is negative in N2, indicating that the output distribution in this treatment could be explained by participants choosing actions with lower expected payoffs.

Figure 11 illustrates the goodness of fit by comparing the kernel density estimation of experimental data to the choice probabilities predicted by the best-fitting QRE. In the standard treatments, QRE can explain the shift in the choice distribution that results from a different market size. In the normalized treatments, the fit is less good as QRE underestimates the frequency of focal outcomes, such as the Nash equilibrium and the endpoints of the strategy space.

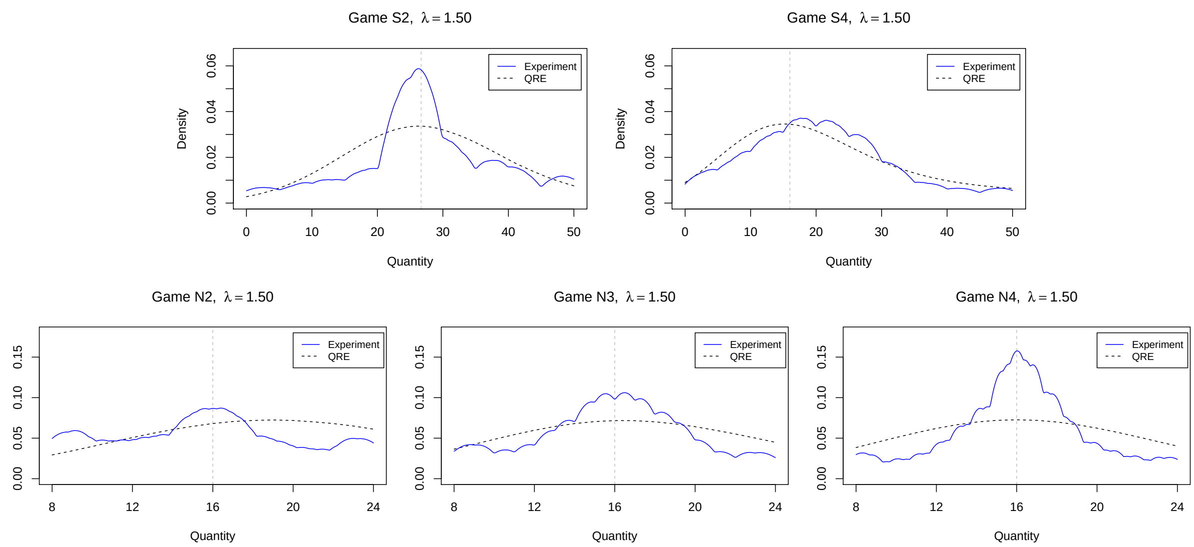


Figure 12: Best-fitting QRE estimated for all treatments (compared to kernel density).

The additional requirement for the noise parameter to be constant across treatments in the second method hardly changes the fit in the standard treatments, but decreases the fit in the normalized treatments, especially in N2 and N4. Figure 12 shows the choice probabilities estimated by a model with a single value of $\hat{\lambda} = 1.497$. The fit remains good in the standard treatments, but QRE fails to predict any difference in the distribution of choices among the three normalized treatments.

Overall, QRE can explain why output is on aggregate more collusive in smaller markets in standard treatments and why there is no difference in the normalized treatments (result 1). QRE can also explain the overall change in the shape of the choice distribution in standard treatments, but it cannot explain why variance decreases in market size in the normalized treatments (result 2).

5.2 Frequent Response Equilibrium

The finding that variance is decreasing in market size only in the normalized treatments could be explained by the difference in how the output of the opponents is aggregated.

In the standard treatments, the *sum* of opponents' output is mapped into the same best-response, regardless of the market size. In the normalized treatments, the *average* of opponents' output is mapped into the same best-response, regardless of the market size. An increase in market size increases the variance of the sum of output chosen by the other firms, but decreases the variance of the average output. Therefore, for a given best-response correspondence, the variance of the best-response distribution would increase in market size in the standard treatments, but decrease in normalized treatments (for more details and an illustration, see Appendix B).

We can test these predictions by comparing the incentives to respond to the observed feedback in experiments. First, we test the prediction that the variance of total output of the opponents is increasing in market size, but the variance of average output is decreasing. For each participant, we calculate the standard deviation of either the total or the mean opponents' output across all 20 rounds and then compute the standard deviation in each treatment. As predicted, the standard deviation of the total opponents' output increases from 10 in S2 to 19 in S4 in standard treatments. In normalized treatments, the standard deviation of the average opponents' output decreases from 4.2 in N2 to 2.3 in N3 and 1.9 in N4.

These differences affect the shape of the best-response distribution. Figure 13 shows the distribution of best responses to the output of the opponents observed in that round, aggregated across all participants and all rounds. These distributions of ex-post rational output would have been observed if participants always had correct beliefs and optimally responded to them. Standard deviation of the best-response distribution is increasing in market size in the standard treatments (5.7 in S2 and 8.1 in S4) but decreasing in the normalized treatments (5.8 in N2, 4.4 in N3 and 3.1 in N4). This result shows that the observation of variance decreasing in market size only in the normalized treatments could be explained by participants responding to observed feedback.

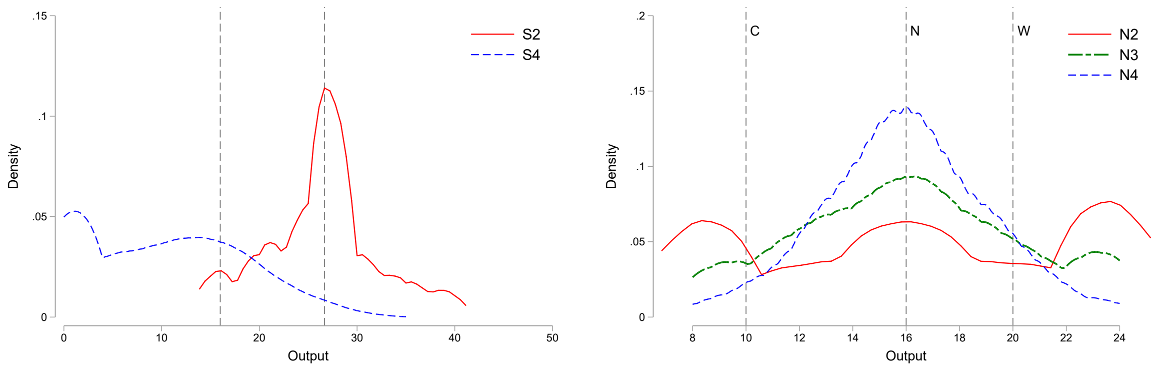


Figure 13: Kernel density of ex-post rational actions, pooled across all rounds.

We use the results about the difference in observed feedback to propose a modification of QRE, which we will refer to as the Frequent Response Equilibrium (FRE). QRE assumes that each player chooses a noisy best response to the choices that everyone else is expected to make; therefore, the choice probabilities are a function of expected payoffs. FRE assumes that each player (noisily) chooses the action that is most likely to be the best response, so the choice probabilities are a function of the expected likelihood to be the best response.[19] QRE could emerge as the long-run outcome of logit response dynamics (Alós-Ferrer and Netzer, 2010; Cason et al., 2021), as players form beliefs from observed history and choose stochastic best-responses. Instead, FRE could emerge as the long-run outcome if players favor actions that are frequent best-responses to the observed history, in a manner similar to probability matching (Vulkan, 2000). It has been shown that when decisions are made from experience, participants are not very sensitive to the average payoff an action generates, but are sensitive to how frequently an action provides the highest payoff (Erev and Barron,

[19]There is a difference between the action that generates the highest expected payoff and an action that is most likely to provide the highest payoff, because the former does not take the magnitude of payoffs into account. For example, consider our treatment N2 and suppose that players expect the opponent to draw their action from a uniform distribution. Compare the attractiveness of the equilibrium output of 16 to the lowest possible output of 8. Producing 16 provides a higher expected payoff than 8 because it generates high profits when the opponent is choosing a low or an intermediate output level (Table E.1). However, 16 is rarely the exact best-response – it is the best response only if the other player chooses 16 as well, whereas 8 is the best response to any output above 21. When evaluated based on the likelihood to be the best response, output of 8 would therefore outperform 16, while 16 would outperform 8 based on the expected payoff.

2005; Yechiam and Busemeyer, 2006).

We wanted FRE to differ from QRE only in the way the attractions are determined (based on the likelihood of being a best-response rather than expected payoff), therefore we retained the assumption that attractions are mapped into choice probabilities using a softmax function with a λ parameter. If $\lambda \to \infty$, FRE approaches Nash equilibrium because the action that has the highest likelihood to be the best response must also provide the highest expected payoff.

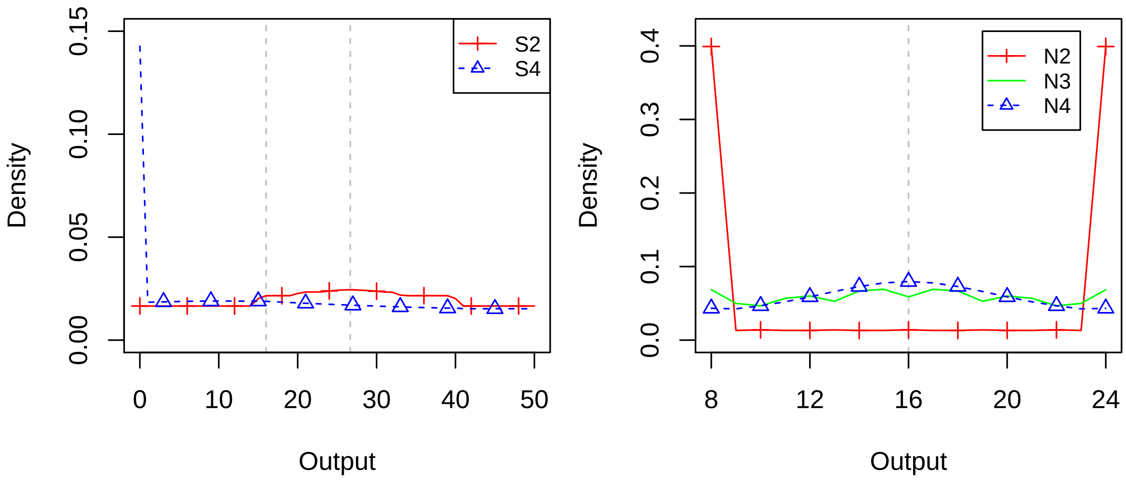
We define FRE formally by extending the definition of QRE from section 3. Let $q_{-i} = \{q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_n\}$ be the pure strategy profile of all players other than i. Then $p_{-i} \in \Delta_{-i}$ is the mixed strategy profile of all others players, where Δ_{-i} is the the set of all possible mixed strategy profiles. The likelihood that q_{-i} will be played in p_{-i} is $p_{-i}(q_{-i})$. Denote the set of strategies that are best responses to a pure strategy profile q_{-i} by $B(q_{-i}) = \{q_k \in Q_i : \forall q_j \in Q_i : \pi_i(q_k, q_{-i}) \geqslant \pi_i(q_j, q_{-i})\}$. Then the likelihood that q_k is the best-response to q_{-i} is calculated by $b(q_k, q_{-i})$, defined as:

$$b(q_k, q_{-i}) = \begin{cases} \frac{1}{|B(q_{-i})|} & \text{if } q_k \in B(q_{-i}) \\ 0 & \text{if } q_k \notin B(q_{-i}) \end{cases}$$

The likelihood that action q_k is the best-response conditional on probabilistic belief p_{-i} is calculated by $r(q_k, p_{-i}) = \sum_{q_{-i}} b(q_k, q_{-i}) p_{-i}(q_{-i})$. FRE assumes that player are maximizing the probability that the action will be the best response; therefore, the utility function defined in equation (2) is replaced by:

$$u_i(q_k, p_{-i}) = r(q_k, p_{-i}) + \varepsilon_{ik} \tag{4}$$

The definition of the solution concept follows the one of QRE, detailed in section 3, with the choice probabilities determined by equation (3). We evaluate the fit of FRE and



(a) Standard treatments (b) Normalized treatments

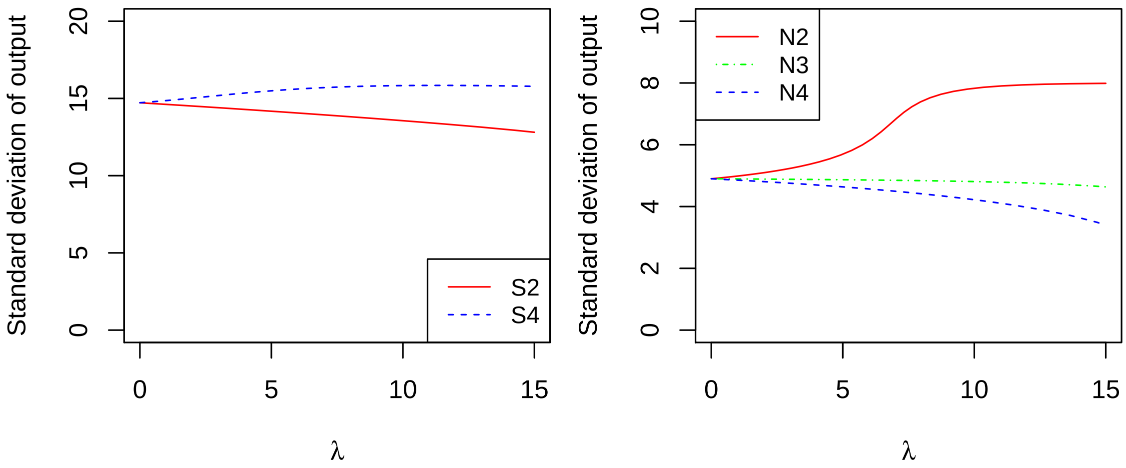
Figure 14: FRE distribution with $\lambda = 8$. Vertical dashed lines indicate symmetric Nash equilibria.

compare it to QRE by calculating the FRE choice distribution at various values of λ and then evaluating the goodness of fit by fitting FRE to the experimental data.

FRE is calculated using the same tracing procedure used for QRE, but instead of using the game's payoffs $\pi_i(q_k, q_{-i})$, we use the likelihood that an action is a best response, calculated by $b(q_k, q_{-i})$. Note that this modification leads to a significant loss of information about the incentives faced by the players.[20]

Figure 14 shows the FRE choice probabilities for $\lambda = 8$. In standard treatments, S4 stands out due to the high predicted frequency of producing nothing, which is the best response when the total output of the other firms exceeds 80. In normalized treatments, the FRE choice distribution has a higher variance in smaller markets. N2 is notable due to a high predicted frequency of the two most extreme output levels. This prediction is explained by a higher likelihood of observing extreme average output in N2 than in N3 or N4, and a best-response function that makes the output of 8 optimal when the average

[20]It should be possible to develop a hybrid model of QRE and FRE, in which attractions are a convex combination of the monetary payoff and the best-response likelihood. However, such a model is beyond the scope of this paper.



(a) Standard treatments (b) Normalized treatments

Figure 15: Standard deviation of output in estimated FRE distributions for $\lambda \in [0, 15]$.

output of the other firms exceeds 21, while 24 is optimal when it falls below 11. Overall, the direction of change in FRE choice distributions in normalized treatments reflects the empirical pattern.

We further explore the differences in variance across market sizes by calculating the standard deviation of the choice distribution for λ values between 0 and 15. Figure 15 shows that FRE correctly predicts higher standard deviation in smaller markets for all λ values in this range, but only in the normalized treatments.

Next, we fit FRE to the data using the method originally developed by Bajari and Hortacsu (2005). First, for each action, we calculate the expected likelihood of being the best response, assuming that the strategy profile of other players is generated by each player independently drawing their strategy from the empirically observed output distribution. The expected likelihoods are mapped into choice probabilities using the softmax function with parameter λ. We fit the model by estimating the value of λ that maximizes the likelihood of the empirical choice distribution.[21]

[21] Note that the value of λ estimated in FRE cannot be compared to the λ values in QRE because the attractions in FRE are measured in likelihood (ranging from 0 to 1) while attractions in QRE represent the expected monetary earnings.

Table 5: Goodness of fit and estimated parameter values in QRE and FRE.

	QRE				FRE			
	Separate		Combined		Separate		Combined	
	$\hat{\lambda}$	LL	$\hat{\lambda}$	LL	$\hat{\lambda}$	LL	$\hat{\lambda}$	LL
S2	1.395	-2739.97	1.497	-2740.34	13.63	-2665.76	7.84	-2701.296
S4	1.535	-4561.56	1.497	-4561.65	4.42	-4701.504	7.84	-4715.934
N2	-0.292	-5436.98	1.497	-5531.25	4.98	-5324.712	7.84	-5369.582
N3	2.507	-4014.05	1.497	-4024.05	14.68	-4001.647	7.84	-4018.677
N4	4.686	-5161.69	1.497	-5281.54	15.32	-5045.288	7.84	-5136.418
Total		-21914.25		-22138.83		-21738.91	7.84	-21941.91

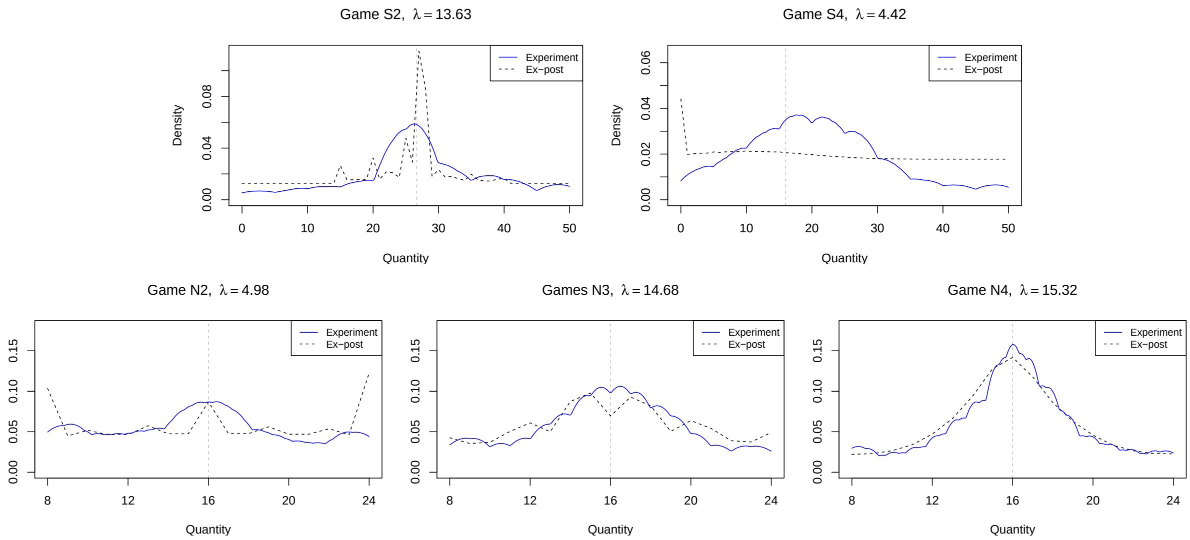


Figure 16: Best-fitting FRE estimated separately for each treatment.

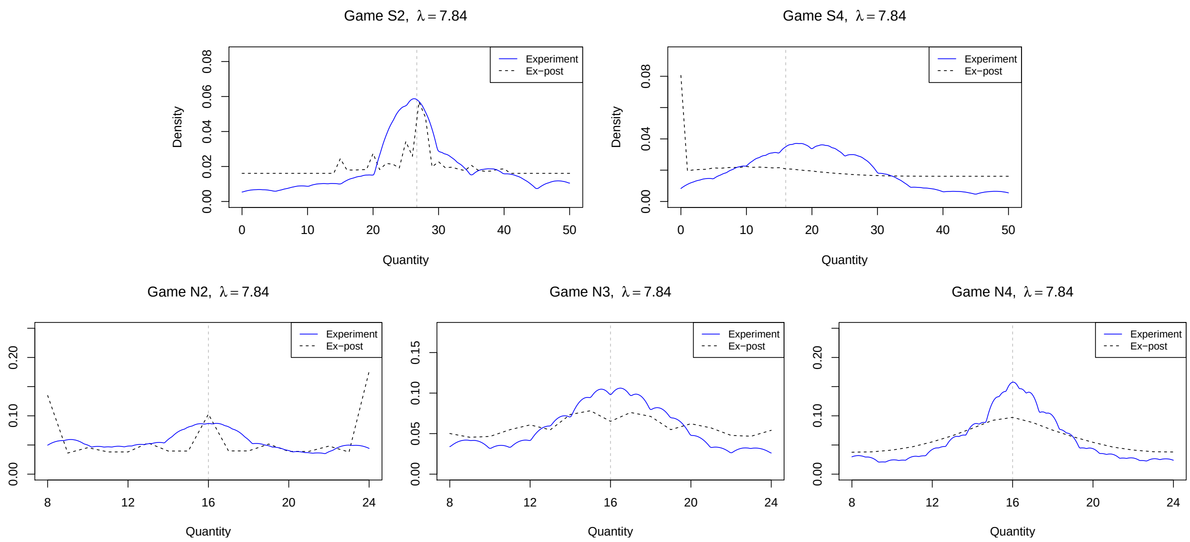


Figure 17: Best-fitting FRE estimated for all treatments.

Table 5 extends the results from Table 4, comparing the estimated values and goodness of fit for QRE and FRE. Figures 16 and 17 illustrate the fit by comparing the choice probabilities in the best-fitting FRE to the kernel density estimates of the experimental data. Overall, FRE has a higher total log-likelihood than QRE, both when a separate model is estimated for each treatment and when a single model is estimated for all treatments. FRE also fits better in each individual game, except for S4. FRE fails to explain choices in S4 because it overestimates the frequency of choosing 0. In the experiment, participants would have maximized their round earnings by producing 0 each time the total output by other participants exceeded 80 (which happened in over 20% of the rounds), yet 0 was chosen only about 2% of the time. In the normalized treatments, predicted choice probabilities are close to the empirical data, although the estimated λ value is much lower in N2 than in N3 or N4. The higher estimated level of noise in N2 is again caused by FRE overestimating the frequency of the most extreme output levels. When λ is required to be the same in all five games, FRE can explain the increased variance in smaller markets of the normalized treatments, although the higher noise level needed to reduce the predicted frequency of the extreme choices in S4 and N2 reduces the goodness of fit in the other three treatments.

Overall, we conclude that the increased variance in smaller markets, found in the normalized treatments, can be explained by a static solution concept based on the frequency with which each action is the best response. This model also provides a better fit than QRE, although it overestimates the frequency of extreme output levels in N2 and S4.

6 Concluding Remarks

Previous Cournot oligopoly experiments found higher rates of tacit collusion in smaller markets. We demonstrate that this result could be explained by changes in the payoff structure that occur when the market size is manipulated using the standard Cournot payoff function. We propose an alternative way to manipulate the market size, which makes markets with a different number of competitors more comparable. This normalized design is used to identify whether differences in collusion rates are driven by the market size itself, or by the changes in incentives or the focality of key outcomes. We replicate the finding of more collusive output in smaller markets using the standard treatments, but find no effect of market size in the normalized treatments. These results suggest that the number effect is largely driven by the changes in incentives or focality, instead of purely the number of interacting firms.

In the normalized treatments, we also find that the variance of individual output is decreasing in market size. This effect cannot be explained by the Quantal Response Equilibrium. Instead, it could be explained by the difference in feedback caused by the implemented aggregation rule: total output in standard treatments, average output in normalized treatments. The variance of average output decreases when the market size goes up, making extreme responses more costly. Instead, the variance of total output increases in larger markets. This difference has consequences beyond the average rates of collusion. Interestingly, increased variance in smaller markets in the normalized treatments provides some support for the original result of higher frequency of collusion in smaller markets, although the frequency of competitive choices goes up as well.

A better understanding of how the number of competitors affects collusion rates could improve the design of competition policy. When deciding whether to approve a merger, regulators evaluate whether the resulting increase in market concentration would significantly lessen competition. The practices used by the Department of Justice and the Federal Trade Commission, two agencies in charge of enforcing antitrust law in the U.S., are explained in the Horizontal Merger Guidelines.[22] The guidelines identify two channels through which

[22]U.S. Department of Justice & Federal Trade Commission, Horizontal Merger Guidelines (2010), avail-

mergers could enhance market power: "unilateral effects" and "coordinated effects".[23] Unilateral effects refer to the higher incentives to increase prices in more concentrated markets; for example, a merger of firms that sell similar products increases the incentive to raise prices because the lost sales that would have been diverted to the competitor's products are now diverted to the products sold by other divisions of the same firm. Coordinated effects refer to the increased likelihood of implicit or explicit coordination on higher prices in more concentrated markets. Successful coordination requires the ability to detect and punish the firms that deviate from collusive agreements,[24] which is easier when there are fewer firms and the behavior of rivals is more predictable. These two channels correspond to the two mechanisms studied in this paper: individual incentives and the pure number effect. Just as the real firms, participants in Cournot oligopoly could collude more in smaller markets because of higher incentives or because implicit coordination is easier when there are fewer participants. Identifying the mechanism is critical to select the appropriate strategy for regulating mergers. If collusion is driven primarily by coordinated effects, regulators should focus on market concentration, as was advocated in the 1968 and 1982 guidelines (Shapiro, 2010). But if, instead, it is primarily driven by the unilateral effects, regulators should instead estimate the value of diverted sales by evaluating the degree of product differentiation, market elasticity of demand or costs of output suppression. Recent versions of the guidelines advocate this view, introducing the concept of unilateral effects in the 1992 guidelines and accentuating it in 2010 (Shapiro, 2010). Consequently, the guidelines have focused more on the economic factors and techniques to estimate the value of diverted sales rather than the market concentration.[25] The results of our experiments support this

able at http://www.justice.gov/atr/public/guidelines/hmg-2010.pdf

[23]The European Commission provides similar guidelines, naming the two channels "coordinated effects" and "non-coordinated effects", see Guidelines on the Assessment of Horizontal Mergers under the Council Regulation on the Control of Concentrations between Undertakings, 2004 O.J. (L 24) 1 (EC), available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2004:031:0005:0018:EN:PDF

[24]U.S. Department of Justice & Federal Trade Commission, Commentary on the Horizontal Merger Guidelines (2006), available at http://www.justice.gov/atr/public/guidelines/215247.pdf

[25]Section 4 of the 2010 Guidelines: "The measurement of market shares and market concentration is

shift, providing evidence that collusion in small markets occurs not because of a smaller number of competitors, but because of the increased incentives to collude. Mergers that do not create additional incentives to collude (e.g. in markets with a low diversion ratio and low margins, Farrell and Shapiro, 2010) might not need to be blocked even if they increase market concentration.

Our study has several limitations that would be interesting to address in future research. We focused only on implicit collusion, but it would be interesting to compare our results to a framework in which explicit collusion is possible, perhaps by adding communication. On one hand, coordination problem might be more important when communication is possible; on the other hand, evidence shows that with communication, even large groups manage to collude, therefore the number effect is not found even with the standard Cournot payoff function (Waichman et al., 2014; Fonseca et al., 2018). It would also be interesting to extend the setup to other games, for example, Bertrand competition, where a similar number effect has been observed for both tacit and explicit collusion (Fonseca and Normann, 2012).

---

not an end in itself, but is useful to the extent it illuminates the merger's likely competitive effects".

References

Alós-Ferrer, C. and Netzer, N. (2010). The logit-response dynamics. *Games and Economic Behavior*, 68(2):413–427.

Bajari, P. and Hortacsu, A. (2005). Are structural estimates of auction models reasonable? evidence from experimental data. *Journal of Political economy*, 113(4):703–741.

Barcelo, H. and Capraro, V. (2015). Group size effect on cooperation in one-shot social dilemmas. *Scientific Reports*, 5(1):1–8.

Bigoni, M. and Fort, M. (2013). Information and learning in oligopoly: An experiment. *Games and Economic Behavior*, 81:192–214.

Bosch-Domènech, A. and Vriend, N. J. (2003). Imitation of successful behaviour in cournot markets. *The Economic Journal*, 113(487):495–524.

Cason, T. N., Friedman, D., and Hopkins, E. (2021). An experimental investigation of price dispersion and cycles. *Journal of Political Economy*, 129(3):789–841.

Chang, C.-C. and Liu, H.-H. (2008). Which is the compromise option? information format and task format as determinants. *Journal of Behavioral Decision Making*, 21(1):59–75.

Choi, S., Goyal, S., and Moisan, F. (2020). Large scale experiments on networks: A new platform with applications. Technical report, Faculty of Economics, University of Cambridge.

Crosetto, P., Weisel, O., and Winter, F. (2019). A flexible z-Tree and oTree implementation of the social value orientation slider measure. *Journal of Behavioral and Experimental Finance*, 23:46–53.

Davies, S., Olczak, M., and Coles, H. (2011). Tacit collusion, firm asymmetries and numbers: evidence from ec merger cases. *International Journal of Industrial Organization*, 29(2):221–231.

Davis, D. (2009). Pure numbers effects, market power, and tacit collusion in posted offer markets. *Journal of Economic Behavior & Organization*, 72(1):475–488.

Diederich, J., Goeschl, T., and Waichman, I. (2016). Group size and the (in) efficiency of pure public good provision. *European Economic Review*, 85:272–287.

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

Engel, C. (2007). How much collusion? a meta-analysis of oligopoly experiments. *Journal of Competition Law & Economics*, 3(4):491–549.

Erev, I. and Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, 112(4):912.

Farrell, J. and Shapiro, C. (1990). Horizontal mergers: an equilibrium analysis. *The American Economic Review*, 80(1):107–126.

Farrell, J. and Shapiro, C. (2010). Antitrust evaluation of horizontal mergers: An economic alternative to market definition. *The BE Journal of Theoretical Economics*, 10(1).

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fonseca, M. A., Li, Y., and Normann, H.-T. (2018). Why factors facilitating collusion may not predict cartel occurrence—experimental evidence. *Southern Economic Journal*, 85(1):255–275.

Fonseca, M. A. and Normann, H.-T. (2012). Explicit vs. tacit collusion—the impact of communication in oligopoly experiments. *European Economic Review*, 56(8):1759–1772.

Fouraker, L. and Siegel, S. (1963). *Bargaining Behavior*. McGraw-Hill.

Friedman, D., Huck, S., Oprea, R., and Weidenholzer, S. (2015). From imitation to collusion: Long-run learning in a low-information environment. *Journal of Economic Theory*, 155:185–205.

Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. *The Review of Economic Studies*, 38(1):1–12.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Hommes, C., Kopányi-Peuker, A., and Sonnemans, J. (2020). Bubbles, crashes and information contagion in large-group asset market experiments. *Experimental Economics*, pages 1–20.

Horstmann, N., Krämer, J., and Schnurr, D. (2018). Number effects and tacit collusion in experimental oligopolies. *The Journal of Industrial Economics*, 66(3):650–700.

Huck, S., Normann, H.-T., and Oechssler, J. (2004). Two are few and four are many: number effects in experimental oligopolies. *Journal of Economic Behavior & Organization*, 53(4):435–446.

Isaac, R. M. and Walker, J. M. (1988). Group size effects in public goods provision: The voluntary contributions mechanism. *The Quarterly Journal of Economics*, 103(1):179–199.

Isaac, R. M., Walker, J. M., and Williams, A. W. (1994). Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups. *Journal of public Economics*, 54(1):1–36.

Ivaldi, M., Jullien, B., Rey, P., Seabright, P., and Tirole, J. (2007). The economics of tacit collusion: Implications for merger control. In *The Political Economy of Antitrust*, pages 217–239. Emerald Group Publishing Ltd.

Jaffe, S. and Weyl, E. G. (2013). The first-order approach to merger analysis. *American Economic Journal: Microeconomics*, 5(4):188–218.

Le Coq, C. and Orzen, H. (2006). Do forward markets enhance competition?: Experimental evidence. *Journal of Economic Behavior & Organization*, 61(3):415–431.

Levenstein, M. C. and Suslow, V. Y. (2006). What determines cartel success? *Journal of Economic Literature*, 44(1):43–95.

Lim, W., Matros, A., and Turocy, T. L. (2014). Bounded rationality and group size in tullock contests: Experimental evidence. *Journal of Economic Behavior & Organization*, 99:155–167.

Masiliūnas, A. (2017). Overcoming coordination failure in a critical mass game: strategic motives and action disclosure. *Journal of Economic Behavior & Organization*, 139:214–251.

Masiliūnas, A. and Nax, H. H. (2020). Framing and repeated competition. *Games and Economic Behavior*, 124:604–619.

McFadden, D. (1981). Econometric models of probabilistic choice. In Manski, C. and McFadden, D., editors, *Structural Analysis of Discrete Data with Econometric Applications*. Harvard University Press.

McKelvey, R. D., McLennan, A. M., and Turocy, T. L. (2015). Gambit: Software tools for game theory.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

Nosenzo, D., Quercia, S., and Sefton, M. (2015). Cooperation in small groups: the effect of group size. *Experimental Economics*, 18(1):4–14.

Oechssler, J., Roomets, A., and Roth, S. (2016). From imitation to collusion: a replication. *Journal of the Economic Science Association*, 2(1):13–21.

Orzen, H. (2008). Counterintuitive number effects in experimental oligopolies. *Experimental Economics*, 11(4):390–401.

Raven, J. C. and Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales.* Oxford pyschologists Press.

Roux, C. and Thöni, C. (2015). Collusion among many firms: The disciplinary power of targeted punishment. *Journal of Economic Behavior & Organization*, 116:83–93.

Shapiro, C. (2010). The 2010 horizontal merger guidelines: From hedgehog to fox in forty years. *Antitrust Law Journal*, 77(1):49–107.

Suetens, S. and Potters, J. (2007). Bertrand colludes more than cournot. *Experimental Economics*, 10(1):71–77.

Turocy, T. L. (2005). A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence. *Games and Economic Behavior*, 51(2):243–263.

Valenzuela, A. and Raghubir, P. (2009). Position-based beliefs: The center-stage effect. *Journal of Consumer Psychology*, 19(2):185–196.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, 80(1):234–248.

Van Koten, S. and Ortmann, A. (2013). Structural versus behavioral remedies in the deregulation of electricity markets: An experimental investigation motivated by policy concerns. *European Economic Review*, 64:256–265.

Vega-Redondo, F. et al. (1997). The evolution of walrasian behavior. *Econometrica*, 65(2):375–384.

Vives, X. (1984). Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics*, 15(4):546–554.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, 14(1):101–118.

Waichman, I., Requate, T., et al. (2014). Communication in cournot competition: An experimental study. *Journal of Economic Psychology*, 42:1–16.

Yechiam, E. and Busemeyer, J. R. (2006). The effect of foregone payoffs on underweighting small probability events. *Journal of Behavioral Decision Making*, 19(1):1–16.

Zelmer, J. (2003). Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6(3):299–310.

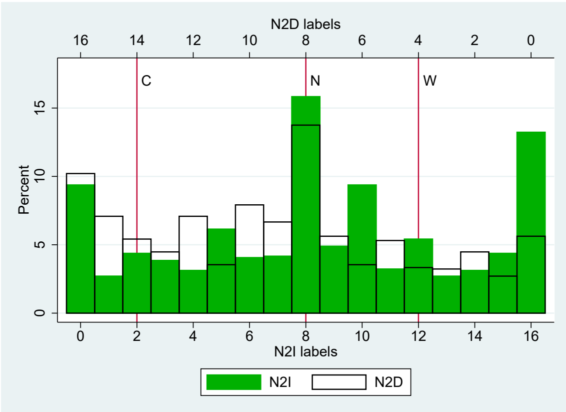
Appendix

A Strategy Space Labeling

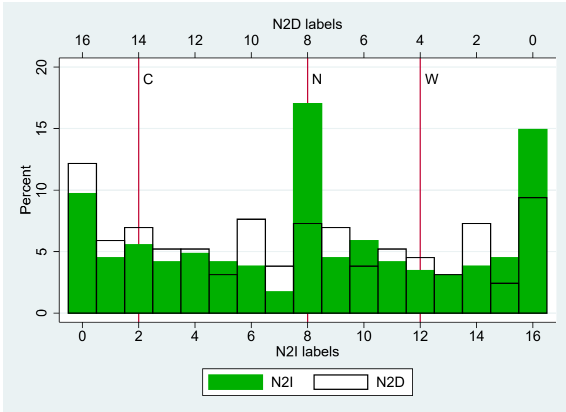
Figure A.18 shows the distribution of choices in the normalized treatment, comparing the two different labeling schemes. In the increasing scheme, quantities from 8 to 24 were labeled from 0 to 24. In the decreasing scheme, the labeling was reversed, with the choice of 0 corresponding to the output of 24 and the choice of 16 corresponding to the output of 8. Figure A.18 shows that labeling has some effect on the distribution of choices. The original labels seen by participants are displayed on the two axes. Data shows that participants make competitive choices more often in the increasing treatments, for all market sizes. In part, this is explained by action "10" being more focal and frequently chosen in all treatments. In increasing treatments, the action labeled as "10" is midway between the Nash and Walrasian equilibria, but in decreasing treatments it is more collusive than the Nash equilibrium prediction. However, there are other differences as well. With all three market sizes, action 16 is chosen more often than action 0, regardless of whether it is mapped into output 8 or 24. There is some evidence that the distribution of choices is shifted towards the actions with higher labels, especially in the 3-firm market. Thus it seems that the effect of labeling is driven both by the focality of action "10" and by a preference for choosing actions labeled with higher numbers.

The right column of Figure A.18 shows the distribution of choices only in the last 5 rounds. The effect of labeling persists in 2 and 3 firm markets, although the magnitude of the effect is lower.

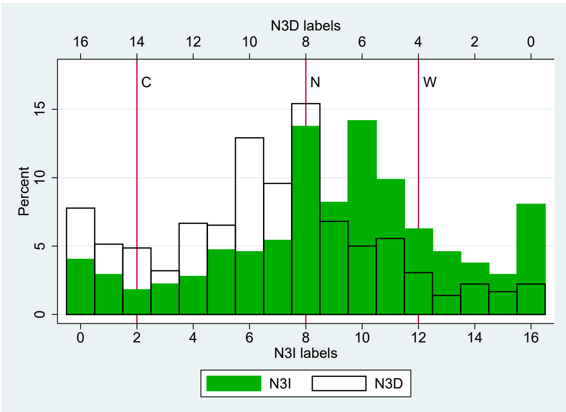
We evaluate the effect of labeling formally by comparing the ratio of average output to Nash equilibrium prediction. Table A.6 shows that on average, the chosen output is above NE prediction in increasing treatments, but below it in decreasing treatments.



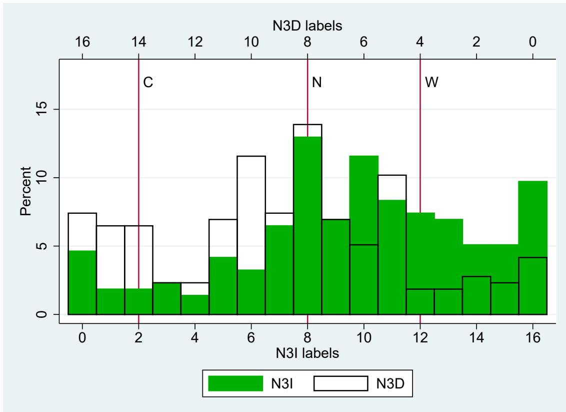
(a) N2, rounds 1-20



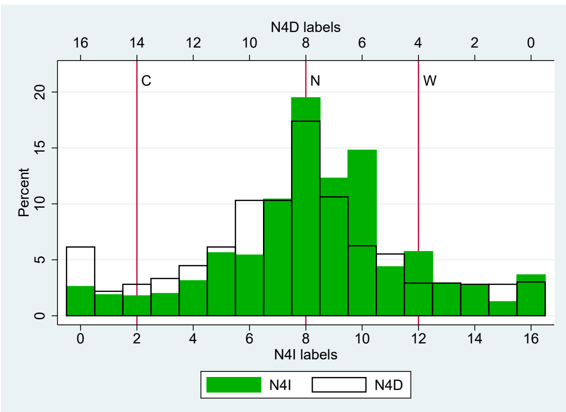
(b) N2, rounds 15-20



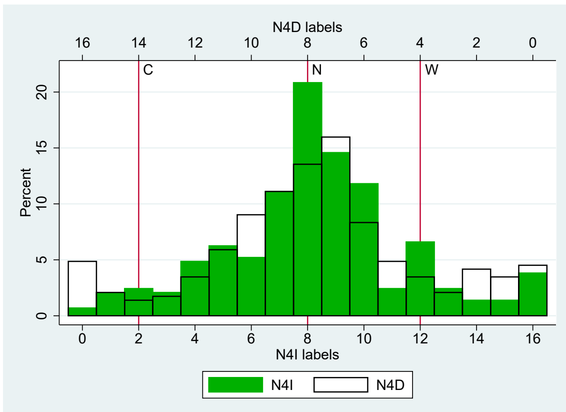
(c) N3, rounds 1-20



(d) N3, rounds 15-20



(e) N4, rounds 1-20



(f) N4, rounds 15-20

Figure A.18: Output distributions in normalized treatments for the increasing and deceasing labeling schemes.

Table A.6: Comparison of the ratio of average output to NE prediction and average profit per round. The number in brackets shows the values averaged across rounds 15-20.

	N2I	N3I	N4I	N2D	N3D	N4D
q_i/q_i^N	1.03	1.07	1.03	0.94	0.92	0.98
	[1.02]	[1.10]	[1.02]	[0.97]	[0.95]	[1.02]
π_i	366.4	322.8	354.2	408.3	419.5	390.6
	[377.5]	[299.5]	[363.4]	[386.3]	[396.6]	[360.3]

Mann-Whitney U test shows that choices are significantly more collusive in the decreasing treatments (MWU $p = 0.0111$ in 4-firm treatments, $p = 0.0085$ in 3-firm treatments and $p = 0.0296$ in 2-firm treatments). If we look only at the last 5 rounds, the difference is significant only in 3-firm market ($p = 0.0178$). A similar result is found in terms of the generated profits, which are significantly higher in decreasing treatments (MWU $p = 0.0153$ in 4-firm treatments, $p = 0.0056$ in 3-firm treatments and $p = 0.0392$ in 2-firm treatments). In the last 5 rounds, the difference is significant only in 3-firm treatments (MWU $p = 0.0243$).

The manipulation of labels allows us to more accurately assess the baseline aggregate levels of collusion. Had we run only the increasing treatments, as is done in all the previous literature, we would likely conclude that there is a tendency to behave more competitively than predicted by the Nash equilibrium, especially in treatments with the market size of 3 and 4. But the comparison to the decreasing treatments reveals that this tendency is driven in part by the labeling rather than the incentive scheme. By combining the data from two different labeling schemes, we can evaluate and eliminate the effect of focal locations.

B Response to Feedback

The finding that variance is decreasing in market size only in the normalized treatments could be explained by the difference in how the output of the opponents is aggregated in the two designs. In the standard treatments, the *sum* of opponents' output is mapped into the same best-response, regardless of the market size. In the normalized treatments, the *average* of opponents' output is mapped into the same best-response, regardless of the market size. The distribution of the total and average output varies with the market size, affecting the variance of the best-response distribution. As an illustration, suppose that each opponent chooses their output q_i from a normal distribution $\mathcal{N}(\mu, \sigma^2)$. Then the sum of the total output chosen by $(n-1)$ opponents is the sum of $(n-1)$ random variables drawn from normal distributions. The probability distribution of the total output is a normal distribution with the following parameters:

$$\sum_{j=1}^{n-1} q_j \sim \mathcal{N}((n-1)\mu, (n-1)\sigma^2)$$

The probability distribution of the average output chosen by $(n-1)$ opponents is:

$$\frac{1}{n-1}\sum_{j=1}^{n-1} q_j \sim \mathcal{N}(\mu, \frac{1}{n-1}\sigma^2)$$

Note that as the market size increases, the variance of the total output goes up, but the variance of the average output goes down. This difference is subsequently translated into a difference in the variance of the best-response distribution, as illustrated in Figures B.1 and B.2. The curves inside the figures display the probability distributions of the payoff-relevant statistics – total output for standard treatments and average output for normalized treatments. The distributions are plotted assuming that the mean of the distribution is equal to the Nash equilibrium and the standard deviation is equal to 8 in standard and 3

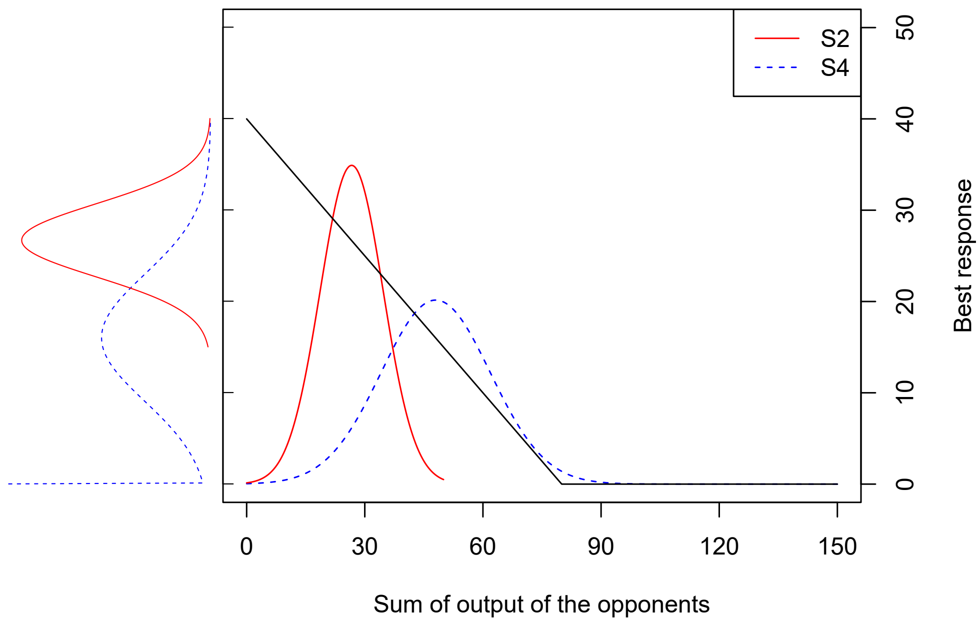


Figure B.1: Best-response curve, probability distribution of average output of the opponents and the resulting probability distribution of the best response (on the left). Standard deviation equals to 8.

in normalized treatments (close to the standard deviations seen in experiments, as shown in Figure 9). The variance of the total opponents' output is increasing in market size (Figure B.1), but the variance of the mean output is decreasing in market size (Figure B.2). The black line in each figure shows how the total or mean output chosen by opponents (which is on the x-axis) is mapped into the best response. The curves on the left side of the figure show the resulting distribution of best-responses to the corresponding distribution of either the sum or the mean output of the opponents. As one can observe, a higher variance of the latter distribution translates into a higher variance of the best-response distribution. We conclude that the variance of the best-response distribution would be predicted to increase in market size in standard treatments, but decrease in market size in normalized treatments.

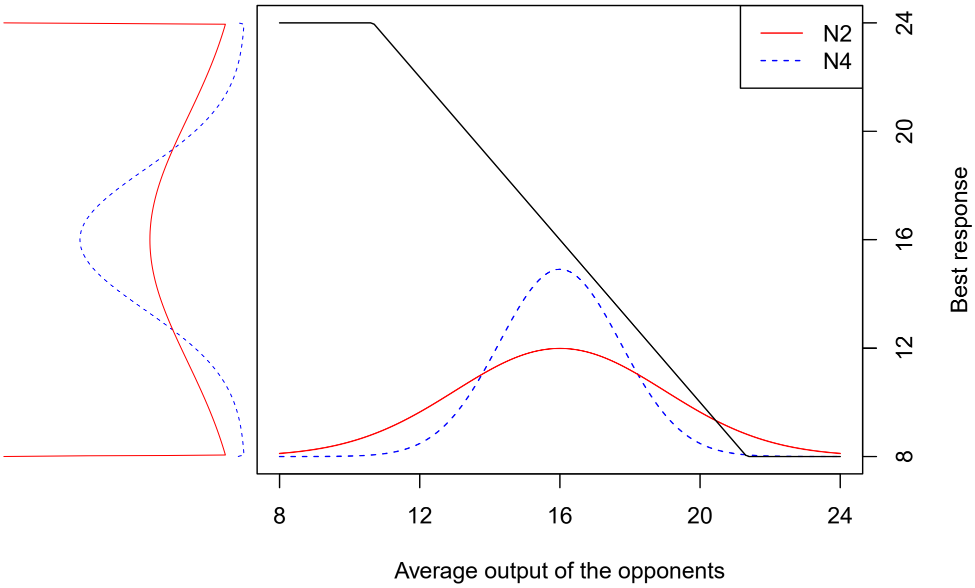


Figure B.2: Best-response curve, probability distribution of average output of the opponents and the resulting probability distribution of the best response (on the left). Standard deviation equals to 3.

C Additional Results

Table C.1: Structure of all treatments that were run. This paper uses data only from the first block of 20 rounds. In games marked with I, the strategy space 8-24 was mapped into 0-16. In games marked with D, the strategy space was reversed before mapping into 0-16. Games marked with S use a standard Cournot oligopoly incentive structure with a strategy space 0-50. SC3 is a 3-person game of strategic complements, with either a strategy space 0-16 (SC3) or 0-50 (SC').

Treatment	1	2	3	4	5	6	7	8
Rounds 1-20	N2I	N4I	N3I	N2D	N4D	N3D	S2	S4
Rounds 21-40	N3I	N3I	SC3	N3D	N3D	SC3	S3	S3
Rounds 41-60	SC3	SC3	-	SC3	SC3	-	SC3'	SC3'
# Sessions	4	4	3	4	4	3	3	5
# Participants	48	48	36	48	48	36	36	60
# Markets in B1	24	12	12	24	12	12	18	15

Table C.2: Random effects GLS regression. Standard errors clustered on the group level. Data from all rounds.

	Increasing			Decreasing		
	DV: r	DV: φ^N	DV: φ^W	DV: r	DV: φ^N	DV: φ^W
3-firm market	0.0412	-0.110	-0.0659	-0.0193	0.0516	0.0309
	(1.61)	(-1.61)	(-1.61)	(-0.44)	(0.44)	(0.44)
4-firm market	0.000326	-0.000868	-0.000521	0.0375	-0.1000	-0.0600
	(0.01)	(-0.01)	(-0.01)	(1.33)	(-1.33)	(-1.33)
1/Round	0.00697	-0.0186	-0.0112	-0.0834***	0.222***	0.133***
	(0.23)	(-0.23)	(-0.23)	(-2.82)	(2.82)	(2.82)
Constant	1.024***	-0.0651	0.361***	0.953***	0.124*	0.475***
	(49.51)	(-1.18)	(10.90)	(34.52)	(1.69)	(10.74)
N	2640	2640	2640	2640	2640	2640

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: Random effects GLS regression. Standard errors clustered on the group level. Data from rounds 15-20.

	Standard			Normalized		
	DV: r	DV: φ^N	DV: φ^W	DV: r	DV: φ^N	DV: φ^W
3-firm market	–	–	–	0.0322	-0.0858	-0.0515
				(0.85)	(-0.85)	(-0.85)
4-firm market	0.182***	-0.418**	-0.571***	0.0256	-0.0683	-0.0410
	(3.79)	(-2.56)	(-8.07)	(0.93)	(-0.93)	(-0.93)
1/Round	1.861	-6.749	-2.619	-1.574	4.197	2.518
	(0.43)	(-0.53)	(-0.39)	(-1.15)	(1.15)	(1.15)
D Treatments				-0.0612**	0.163**	0.0979**
				(-2.40)	(2.40)	(2.40)
Constant	0.943***	0.187	0.750*	1.112***	-0.300	0.220*
	(3.64)	(0.24)	(1.87)	(13.82)	(-1.40)	(1.71)
N	576	576	576	1584	1584	1584

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Goodness of fit and estimated parameter values in QRE and FRE, separately for standard and normalized treatments.

	QRE		FRE	
Treatment	$\hat{\lambda}$	LL	$\hat{\lambda}$	LL
S2	1.482	-2740.24	8.28	-2696.34
S4	1.482	-4561.73	8.28	-4720.376
N2	1.526	-5534.16	7.70	-5365.159
N3	1.526	-4023.46	7.70	-4019.388
N4	1.526	-5279.18	7.70	-5139.97
Total		-22138.77		-21941.23

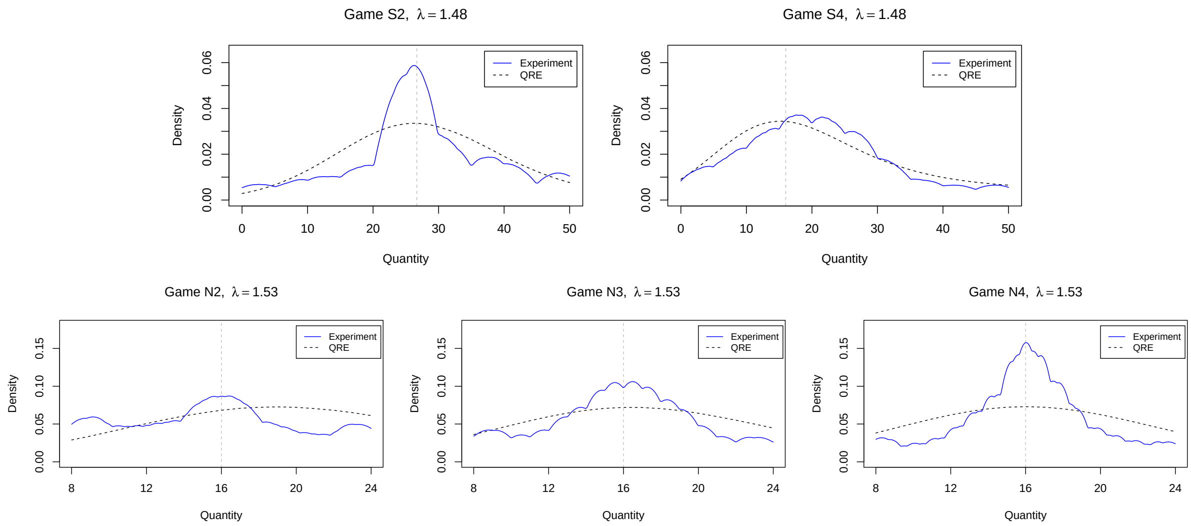


Figure C.1: Best-fitting QRE estimated separately for standard and normalized treatments (compared to kernel density).

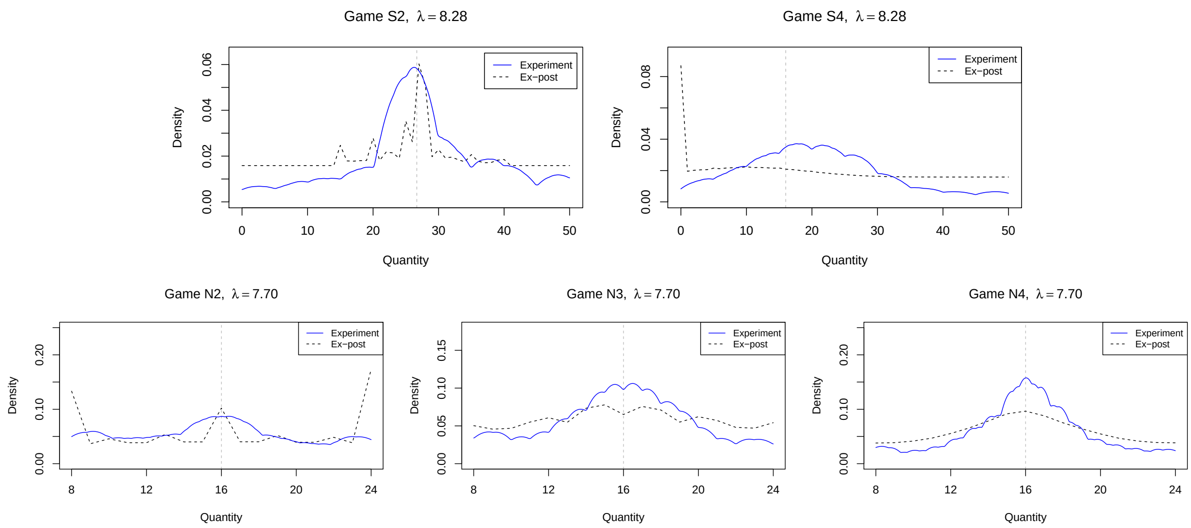


Figure C.2: Best-fitting FRE estimated separately for standard and normalized treatments (compared to kernel density).

D Instructions

The original French version of instructions is available on request. We provide an English translation below. We reproduce only the instructions for the treatments with the "normalized" design. The only difference in the instructions for the "standard" design is the cardinality of the strategy space: number "16" in highlighted places was replaced by "50".

GENERAL INSTRUCTIONS

Welcome to the Laboratory of Experimental Economics of Nice (LEEN – Nice Lab).

By agreeing to participate in this experiment, you agree with the regulations of the laboratory, which are available on our website or on request.

In this experiment your decisions will be anonymous and will partly determine your final payment, therefore read the following instructions carefully. The participation fee of 5 EUR is included in the payoff function. Your earnings will be paid to you individually and confidentially private once you complete a short questionnaire at the end of the experiment.

In this experiment you can earn money. During the experiment we will refer to ECU (Experimental Currency Unit) instead of EUR. The total amount of ECU that you will have earned during the experiment will be converted into cash and paid individually at the end of the experiment. The conversion rate used to convert your ECU into your cash payment will be **150 ECU = 1 EUR**.

We ask you not to communicate or to disturb the other participants. We also ask you to turn off your mobile phones and not use them during the experiment.

In these rules are not followed, the experiment may be stopped and all payments canceled.

If you encounter a technical problem, we ask you to raise your hand silently and wait for the experimenter.

All the participants with whom you interact during this experiment will receive the same instructions and participate in the same experiment.

DESCRIPTION OF THE EXPERIMENT

The experiment will have several parts. Each part will consist of 20 rounds. At the end of the experiment one round from each part will be randomly selected for payment. All rounds have an equal chance to be selected. Your earnings from the selected rounds will be added up, converted into cash and paid to you in private.

In each round you will be matched with other participants. At the start of each part you will be informed about how many participants you will be playing with. You will choose a number (between 0 and 16), which we will call "your action". Every other participant will choose an action at the same time. Your payoff will depend on your action and on the average action of all other participants with whom you were matched. In each round, you will have 30 seconds to make a decision. To make a choice, you must enter your action into the field at the top and click "OK" before the time runs out. The participants with whom you will interact will face the same task as you and will have the same information and payoff function. The task, the payoff function and the participants with whom you will interact will be the same in each round of one part. In each new part, you will play against participants with whom you did not interact in previous parts.

The exact way of how your payoff depends on your action and on the average action of other participants will be explained using a payoff table and a payoff calculator, which will be available on the computer screen when you will be making your decision.

- **The payoff table** shows your payoffs for some combinations of your action and the average action of other participants.

- **The payoff calculator** allows you to enter any action for yourself and an average action of other participants, and displays a payoff that you would receive in that case.

Starting from round 2, you will also be informed about your payoff in the previous round. Furthermore, you will have an option to view the following additional information:

1. **Average choices and their history.** This option gives you information about the average choice of other participants and your payoff in the previous round, as well as in all earlier rounds of that part.

2. **Individual choices and payoffs.** This option gives you information about the choices and payoffs of each member in your group, including yourself.

You will be able to switch between these options using buttons on your computer screen.

In addition, after the first round of each part we will ask you to guess the average action of other participants in that round. The closer your guess is to the average choice of other participants, the higher will be your payment. If your guess is G and the actual average action of other participants is D, your payment will be higher the smaller is the absolute difference between G and D (denoted $|G - D|$). In particular, your payoff will be: $\left(1 - \frac{|G-D|}{16}\right) * 100$ ECU. Notice that if your guess is exactly equal to the average choice ($G - D = 0$), you will receive 100 ECU. At the end of the experiment one of these tasks will be randomly chosen. The payment from the chosen task will be added to your earnings.

At the end of all parts you will be informed about your payoff in ECU from the rounds that were randomly selected for payment. Payoff from these rounds will be summed up, converted into EUR and paid in private once you complete a short questionnaire. In the questionnaire you will have a chance to make additional income which will be added to your earnings. Please stay seated until we ask you to come to receive the earnings.

If you have any further questions, please raise your hand now. The experiment will start once everyone has finished reading the instructions.

E Payoff Tables

Table E.1: Payoffs in N2, N3 and N4 treatments.

	Average output chosen by opponents																
Output	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
8	514	490	466	442	418	394	370	346	322	298	274	250	226	202	178	154	130
9	553	526	499	472	445	418	391	364	337	310	283	256	229	202	175	148	121
10	590	560	530	500	470	440	410	380	350	320	290	260	230	200	170	140	120
11	625	592	559	526	493	460	427	394	361	328	295	262	229	196	163	130	119
12	658	622	586	550	514	478	442	406	370	334	298	262	226	190	154	118	118
13	689	650	611	572	533	494	455	416	377	338	299	260	221	182	143	117	117
14	718	676	634	592	550	508	466	424	382	340	298	256	214	172	130	116	116
15	745	700	655	610	565	520	475	430	385	340	295	250	205	160	115	115	115
16	770	722	674	626	578	530	482	434	386	338	290	242	194	146	114	114	114
17	793	742	691	640	589	538	487	436	385	334	283	232	181	130	113	113	113
18	814	760	706	652	598	544	490	436	382	328	274	220	166	112	112	112	112
19	833	776	719	662	605	548	491	434	377	320	263	206	149	111	111	111	111
20	850	790	730	670	610	550	490	430	370	310	250	190	130	110	110	110	110
21	865	802	739	676	613	550	487	424	361	298	235	172	109	109	109	109	109
22	878	812	746	680	614	548	482	416	350	284	218	152	108	108	108	108	108
23	889	820	751	682	613	544	475	406	337	268	199	130	107	107	107	107	107
24	898	826	754	682	610	538	466	394	322	250	178	106	106	106	106	106	106

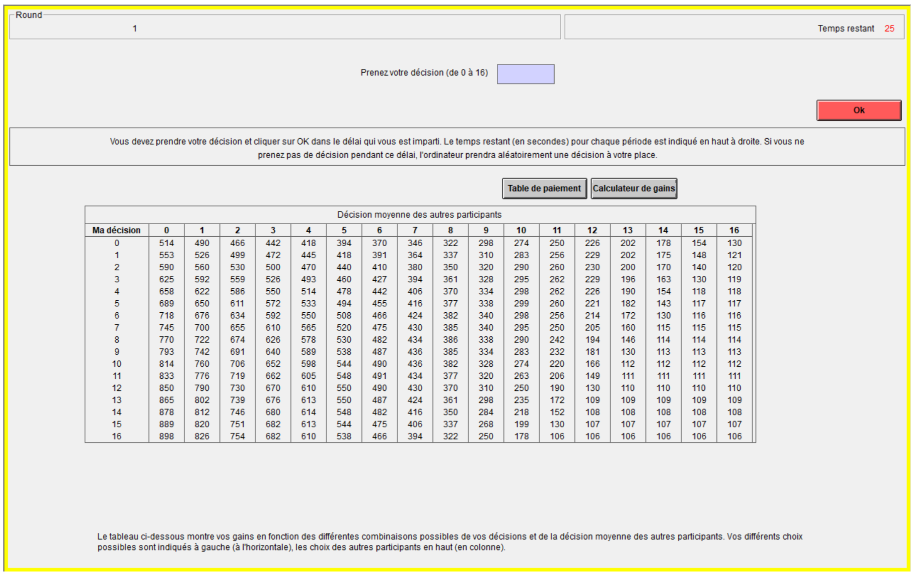
Table E.2: Payoffs in S2 treatment.

	Average output chosen by opponents																				
Output	0	2.5	5.0	7.5	10.0	12.5	15.0	17.5	20.0	22.5	25.0	27.5	30.0	32.5	35.0	37.5	40.0	42.5	45.0	47.5	50.0
0	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130
2.5	200	197	195	193	191	188	186	184	182	179	177	175	173	170	168	166	164	161	159	157	155
5	265	260	256	251	247	242	238	233	229	224	220	215	211	206	202	197	193	188	184	179	175
7.5	326	319	312	305	299	292	285	278	272	265	258	251	245	238	231	224	218	211	204	197	191
10	382	373	364	355	346	337	328	319	310	301	292	283	274	265	256	247	238	229	220	211	202
12.5	434	422	411	400	389	377	366	355	344	332	321	310	299	287	276	265	254	242	231	220	209
15	481	467	454	440	427	413	400	386	373	359	346	332	319	305	292	278	265	251	238	224	211
17.5	524	508	492	476	461	445	429	413	398	382	366	350	335	319	303	287	272	256	240	224	209
20	562	544	526	508	490	472	454	436	418	400	382	364	346	328	310	292	274	256	238	220	202
22.5	596	575	555	535	515	494	474	454	434	413	393	373	353	332	312	292	272	251	231	211	191
25	625	602	580	557	535	512	490	467	445	422	400	377	355	332	310	287	265	242	220	197	175
27.5	650	625	600	575	551	526	501	476	452	427	402	377	353	328	303	278	254	229	204	179	155
30	670	643	616	589	562	535	508	481	454	427	400	373	346	319	292	265	238	211	184	157	130
32.5	686	656	627	598	569	539	510	481	452	422	393	364	335	305	276	247	218	188	159	130	118
35	697	665	634	602	571	539	508	476	445	413	382	350	319	287	256	224	193	161	130	117	117
37.5	704	670	636	602	569	535	501	467	434	400	366	332	299	265	231	197	164	130	117	117	117
40	706	670	634	598	562	526	490	454	418	382	346	310	274	238	202	166	130	116	116	116	116
42.5	704	665	627	589	551	512	474	436	398	359	321	283	245	206	168	130	115	115	115	115	115
45	697	656	616	575	535	494	454	413	373	332	292	251	211	170	130	114	114	114	114	114	114
47.5	686	643	600	557	515	472	429	386	344	301	258	215	173	130	113	113	113	113	113	113	113
50	670	625	580	535	490	445	400	355	310	265	220	175	130	112	112	112	112	112	112	112	112

Table E.3: Payoffs in S4 treatment.

	Average output chosen by opponents																									
Output	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50
0	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130	130
2	286	274	262	250	238	226	214	202	190	178	166	154	142	130	128	128	128	128	128	128	128	128	128	128	128	128
4	434	410	386	362	338	314	290	266	242	218	194	170	146	126	126	126	126	126	126	126	126	126	126	126	126	126
6	574	538	502	466	430	394	358	322	286	250	214	178	142	124	124	124	124	124	124	124	124	124	124	124	124	124
8	706	658	610	562	514	466	418	370	322	274	226	178	130	122	122	122	122	122	122	122	122	122	122	122	122	122
10	830	770	710	650	590	530	470	410	350	290	230	170	120	120	120	120	120	120	120	120	120	120	120	120	120	120
12	946	874	802	730	658	586	514	442	370	298	226	154	118	118	118	118	118	118	118	118	118	118	118	118	118	118
14	1054	970	886	802	718	634	550	466	382	298	214	130	116	116	116	116	116	116	116	116	116	116	116	116	116	116
16	1154	1058	962	866	770	674	578	482	386	290	194	114	114	114	114	114	114	114	114	114	114	114	114	114	114	114
18	1246	1138	1030	922	814	706	598	490	382	274	166	112	112	112	112	112	112	112	112	112	112	112	112	112	112	112
20	1330	1210	1090	970	850	730	610	490	370	250	130	110	110	110	110	110	110	110	110	110	110	110	110	110	110	110
22	1406	1274	1142	1010	878	746	614	482	350	218	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108	108
24	1474	1330	1186	1042	898	754	610	466	322	178	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106
26	1534	1378	1222	1066	910	754	598	442	286	130	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104	104
28	1586	1418	1250	1082	914	746	578	410	242	102	102	102	102	102	102	102	102	102	102	102	102	102	102	102	102	102
30	1630	1450	1270	1090	910	730	550	370	190	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
32	1666	1474	1282	1090	898	706	514	322	130	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98	98
34	1694	1490	1286	1082	878	674	470	266	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96	96
36	1714	1498	1282	1066	850	634	418	202	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
38	1726	1498	1270	1042	814	586	358	130	92	92	92	92	92	92	92	92	92	92	92	92	92	92	92	92	92	92
40	1730	1490	1250	1010	770	530	290	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
42	1726	1474	1222	970	718	466	214	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88
44	1714	1450	1186	922	658	394	130	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86	86
46	1694	1418	1142	866	590	314	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84	84
48	1666	1378	1090	802	514	226	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82	82
50	1630	1330	1030	730	430	130	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80

F Screenshots



Round

1

Temps restant 25

Prenez votre décision (de 0 à 16)

Ok

Vous devez prendre votre décision et cliquer sur OK dans le délai qui vous est imparti. Le temps restant (en secondes) pour chaque période est indiqué en haut à droite. Si vous ne prenez pas de décision pendant ce délai, l'ordinateur prendra aléatoirement une décision à votre place.

Table de paiement | Calculateur de gains

	Décision moyenne des autres participants																
Ma décision	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
0	514	490	466	442	418	394	370	346	322	298	274	250	226	202	178	154	130
1	553	526	499	472	445	418	391	364	337	310	283	256	229	202	175	148	121
2	590	560	530	500	470	440	410	380	350	320	290	260	230	200	170	140	120
3	625	592	559	526	493	460	427	394	361	328	295	262	229	196	163	130	119
4	658	622	586	550	514	478	442	406	370	334	298	262	226	190	154	118	118
5	689	650	611	572	533	494	455	416	377	338	299	260	221	182	143	117	117
6	718	676	634	592	550	508	466	424	382	340	298	256	214	172	130	116	116
7	745	700	655	610	565	520	475	430	385	340	295	250	205	160	115	115	115
8	770	722	674	626	578	530	482	434	386	338	290	242	194	146	114	114	114
9	793	742	691	640	589	538	487	436	385	334	283	232	181	130	113	113	113
10	814	760	706	652	598	544	490	436	382	328	274	220	166	112	112	112	112
11	833	776	719	662	605	548	491	434	377	320	263	206	149	111	111	111	111
12	850	790	730	670	610	550	490	430	370	310	250	190	130	110	110	110	110
13	865	802	739	676	613	550	487	424	361	298	235	172	109	109	109	109	109
14	878	812	746	680	614	548	482	416	350	284	218	152	108	108	108	108	108
15	889	820	751	682	613	544	475	406	337	268	199	130	107	107	107	107	107
16	898	826	754	682	610	538	466	394	322	250	178	106	106	106	106	106	106

Le tableau ci-dessous montre vos gains en fonction des différentes combinaisons possibles de vos décisions et de la décision moyenne des autres participants. Vos différents choix possibles sont indiqués à gauche (à l'horizontale), les choix des autres participants en haut (en colonne).

Figure F.1: Payoff table in normalized treatments.

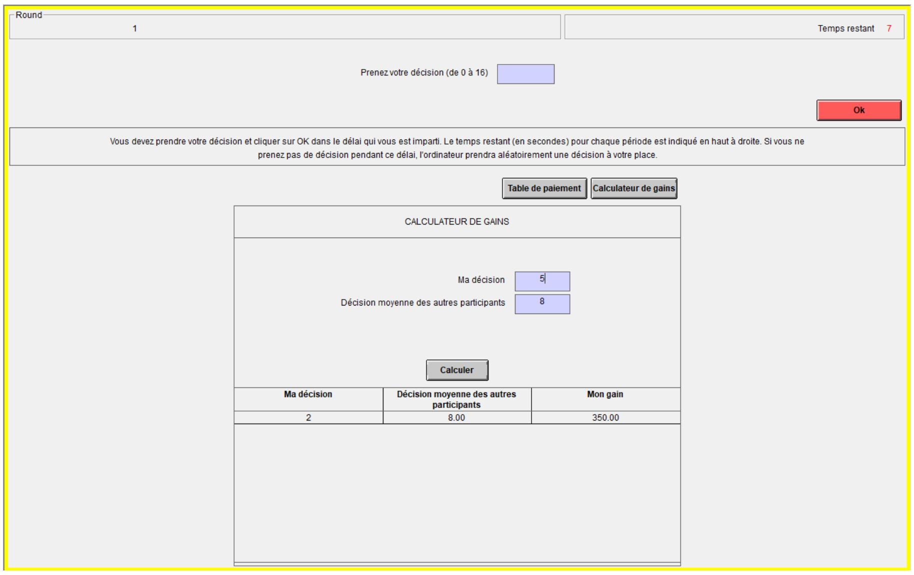


Figure F.2: Payoff calculator in normalized treatments.

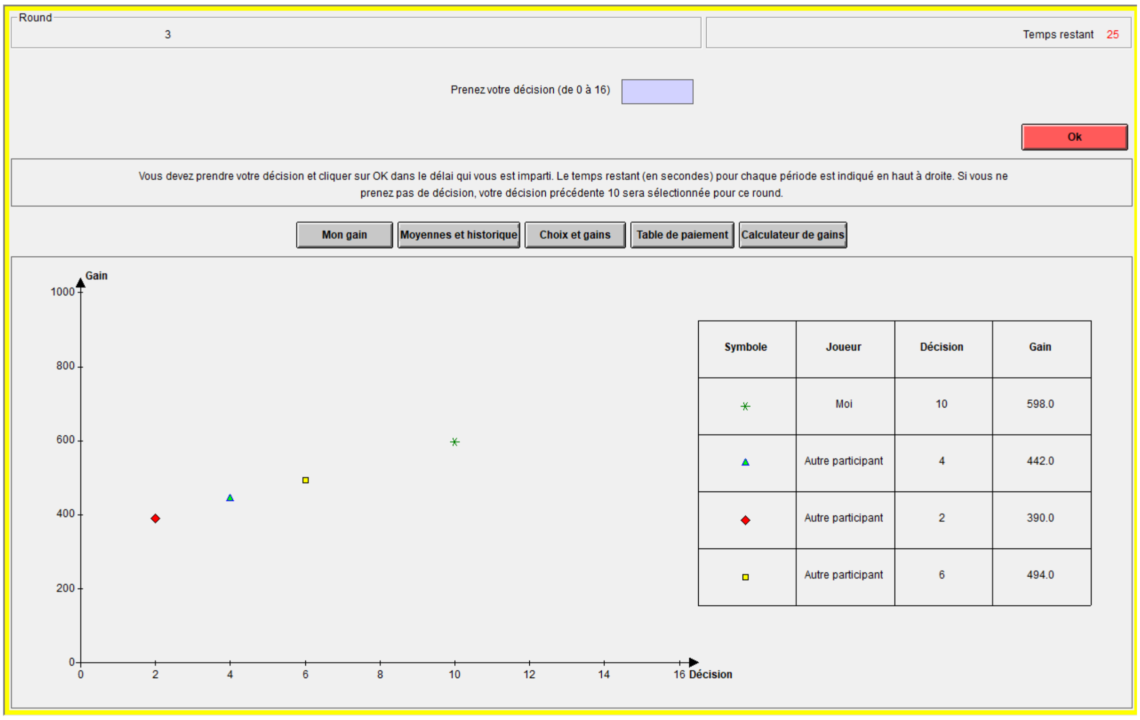


Figure F.3: Feedback about individual choices and payoffs in normalized treatments.

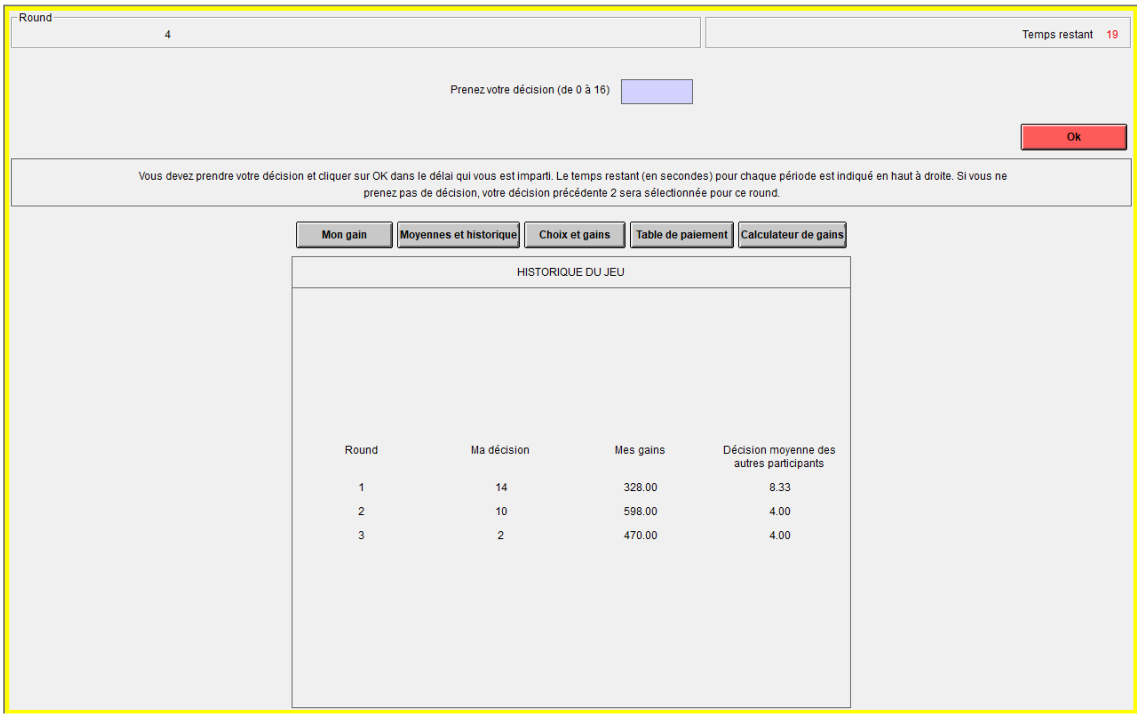


Figure F.4: Feedback about the history of own choices and payoffs in previous rounds.