Efficiency in Queuing Under Decentralized Mechanisms

Kristian López Vargas [*], Brett Williams [†], and Shuchen Zhao[‡]

November 10, 2022

Abstract

In this paper, we study simple mechanisms to reallocate positions in a decentralized queuing problem. In our realistic environment, laboratory subjects "stand" in a virtual queue that moves in real time, waiting to receive some service value. We mainly focus on four decentralized mechanisms: monetary trade, simple swapping, non-monetary token coins, and cheap talk. We find minimal efficiency gains when exchange is excluded, and relatively larger improvements in queues with fiscal, rather than social, currency when exchange is included. In addition, free-text communication has little impact on the queue efficiency.

JEL Classification: C92, D47, D82.

Keywords: Queuing, Decentralized mechanisms, Egalitarian, Auction, Social currency, Communication.

[*]Economics Departments, University of California Santa Cruz; klopezva@ucsc.edu;

[†]AGORA Centre for Market Design, UNSW Sydney; brett.williams2@unsw.edu.au;

[‡]Institute for Advanced Economic Research, Dongbei University of Finance and Economics; fenix.zhaoshuchen@dufe.edu.cn.

1 Introduction

In markets with excess demand, queuing is one of the main methods for firms and governments to allocate scarce resources. This method is widely adopted in housing allocation, access to sports facilities, concert ticketing and many other industries. Understanding and improving queues and their efficiency is an important task for researchers of many fields, including economics, computer science and mathematics.

Mechanism design in relation to this queuing problem faces considerable challenges. In most empirical scenarios, customers are served on a first-come first-serve (FCFS) basis. This approach considers fairness, as participants supposedly arrive according to private values of the item and their own time, but is inefficient compared to socially optimal solutions. While centralized auctions have been exploited to theoretically solve the inefficiency problem (e.g., Kleinrock (1967), Yang et al. (2016)), they are not perceived as simple nor fair by the public audience, compared to the traditional FCFS mechanism.

We study decentralized trading mechanisms with a virtual queue in the laboratory, which are simpler, more intuitive, and therefore more likely to be adopted by humans in empirical situations. In particular, our paper aims at understanding whether these decentralized mechanisms can turn an inefficient queue to an efficient one. Besides the baseline environment where players enter the queue randomly and can only exchange positions with no extra incentives (Swap), our laboratory experiment mainly focuses on three decentralized exchange mechanisms.

We first consider a conventional "take-it-or-leave-it" auction (TL) where the players can make an offer with their exchange request. The theoretical approach lies between the traditional paradigms of bilateral bargaining and multilateral bargaining. Prior to our experiment, we build a descriptive model of the bargaining process in our virtual queue, and implement the same environment as an experimental treatment.

We next consider a mechanism where the players earn non-monetary tokens for their altruistic behavior (Token). The token serves as an additional non-monetary reward and could potentially incentivize the players' egalitarian motivations. In addition, we study how cheap talk affects the queue efficiency when the players can attach a message with their requests under Swap, TL, or Token treatments. This paper contributes to the literature by adding more experimental insights in the decentralized queuing mechanisms, especially when we compare the difference between monetary and non-monetary mechanisms in a virtual queue.

The results are instructive. Compared to the baseline inefficient simple swapping treatment, the take-it-or-leave-it auction turns out to be the most efficient mechanisms but the surplus from the trade goes mostly to the players who own the front positions. On the contrary, the token-coin mechanism and the free-text messaging only weakly improve the egalitarian motivation and does not improve market efficiency. We also find that requesters systematically propose offers that have negative returns in a few exchanges, which can be explained by the joy of winning.

The paper is organized as follows. Section 2 reviews the literature on queuing problems. Section 3 introduces the experimental design. The main experimental results are discussed in section 4. Section 5 concludes the paper with our main findings and future directions.

2 Literature

Because of the importance of queuing, mechanisms have been exploited to increase the efficiency and social welfare in queuing problems. Simple solutions include forbidding entrance and admission fees. Some studies focus on policies that shorten overall waiting time or improve the waiting experience (Kumar et al. (1997), Pruyn and Smidts (1998)). Kleinrock (1967) first shows that a queue's efficiency may be restored if customers' positions depend on how much they pay the server. Though allowing customers who can afford high waiting costs to get ahead can harm the other customers in a queue (Zhou and Soman (2008)), server-based centralized auctions are still one of the most popular solutions.

Regulating the arrival rate is another possible way to improve efficiency, some mechanisms have been developed such as an admission toll and two-part tariff (Edelson and Hilderbrand (1975)) as well as priority price (Mendelson and Whang (1990)). Hassin (1995) applies regulation of the arrival process in a queue with exponential service and without balking by a decentralized self-regulating mechanism under both homogeneous and heterogeneous population and calculated socially optimal arrival rate. Similarly, service rate is another interesting way for regulation (Dewan and Mendelson (1990)).

Balking and queuing information are important factors in queuing, which can make the arrival and staying out decisions an endogenous process. Naor (1969) shows that in queues with balking the individual's decision deviates from the socially preferred one. Rapoport et al. (2004), Seale et al. (2005) and Stein et al. (2007) study this question in a lab environment with decentralized decision making in batch queuing games and find that players move quicker toward equilibrium play when balking is prohibited and when information is public. Meanwhile by aggregate result, behaviour approaches mixed-strategy equilibrium with more experience.

While queuing can cause negative externalities, it can also result in positive externalities via a herding effect. Koo and Fishbach (2010), Giebelhausen et al. (2011) and Kremer and Debo (2012) show that there exists a mental mapping between queue length and quality when quality of products is uncertain. Under this condition, the length of a queue can become a signal about quality, increasing both purchase intention and waiting cost.

Mechanisms promoting trade among agents have been studied in recent years. Kayı and Ramaekers (2010) identify solutions to queuing problems satisfying Pareto-efficiency, equal treatment of equal welfare, symmetry and strategy-proofness by using the 'largest equally distributed pairwise pivotal rule'. In their model, a central server selects all Pareto-efficient queues and sets transfer considering each pair of agents in turn, making each agent in the pair pay the cost she imposes on the pair and distributing the sum of these two payments equal among the others. El Haji and Onderstal (2019) compares the efficiency between a server-initiated auction and customers-initiated trade. They show that the two mechanisms considered do not differ in a statistically meaningful way with respect to the average efficiency gain, irrespective of the arrival protocol. Compared to the previous two mechanisms, Yang et al. (2016) proposes a more efficient trading mechanism in priority queues. The consumers are privately informed about their waiting costs and they mutually agree on the ordering in the queue by trading positions. The paper designs the optimal mechanisms for the social planner, the service provider, and the intermediary who might mediate the trading platform with centralized auctions. In our take-it-leave-it auction treatment, we mainly focus on the decentralized trading mechanism where the players at the back positions can directly propose offers to the players in front of them and exchange positions.

Besides auctions and trades, we also discuss two other decentralized mechanisms: token coins and communication. The idea of token coins raises from Bigoni et al. (2020), where players cooperate in repeated prisoner's dilemma games with non-monetary tokens. In our queuing environment, the token coins are generated from player's altruistic behavior when they exchange backward in the queue and thus represent their reputation. Note that the token coins in our environment serves as a symbol of egalitarianism instead of a alternative monetary system, compared to the one in Bigoni et al. (2020).

Furthermore, cheap talk has been shown as a potential method of improving efficiency and achieving Nash equilibrium (e.g., Farrell and Rabin (1996)). It is also commonly observed in our daily lives when people try to move forward in queues, thus worth our investigation. In our queuing environment, cheap talk could potentially trigger the players' egalitarian motivation and may improve the market efficiency, although the communication itself is not self-committing.

3 Model

3.1 Queues

Our paper rests in a setting defined by a set of queue-goers, $\mathcal{N}$, who have the capacity (and potentially the incentive) to reorganize themselves before service begins. Each queue-goer i, where $i \in \mathcal{N}$, begins in a queue $\mathcal{Q}$, in some position j ($j \in \mathcal{Q}$). We make the following two assumptions for the queue.

1. The length of the queue equals to the number of queue-goers.

2. The queue-goers' initial position in the queue is randomly determined prior to the game.

Each goer i has a private value of service (v_i), drawn from some distribution $F(v)$. We assume goers' values are unique and thus drawn from $F(v)$ without replacement. We make two assumptions with respect to time:

1. service time is standardized, meaning the time associated with each position j is the same, and

2. time is normalized, such that each service takes one unit of time

Thus, the total cost of time (or value lost by waiting longer) for each goer i in position j is $j \cdot v_i$. Alternatively, we can write this in value terms, meaning goer i in position j has a value for the position of $(|\mathcal{N}| - j + 1) \cdot v_i$.

3.2 Messages

Two messaging systems coexist in this environment: (1) one which houses the position exchange requests (to be explained below), and (2) another which allows for information revelation and cheap talk. Here, all messages store and reveal the identity (of the position, rather than the individual) of both the sending queue-goer (s) and receiving queue-goer (r).

Definition 1. *A message, $m(\cdot;\cdot)$, is a tuplet containing a set of information $(s, r, q_r, a; g_r, g_s)$, such that $s < r \in \mathcal{Q}$, $q_r \in \mathbb{Z}_+$, $a \in \{0, 1\}$, and $g_r, g_s \in \mathcal{N}$. This message is sent from some goer in position r, g_r, who is further back in the queue than some goer in position s, g_s, and wishes to switch places. The request may contain some offer amount q_r (per position moved) made by goer g_r. Goer g_s amends and returns the message to goer g_r with the acceptance field, a, filled with either 0 or 1, where 1 denotes an acceptance of the offer and agreement to switch places.*

Henceforth, goer g_r in the above definition will be regarded as a "requester" and goer g_s as a "requestee".

For any message which is accepted, the requester pays the requestee a total of $(r - s) * q_r$, call this $\pi(m)$. Then the problem faced by each goer during the duration of the queue can be written as

$$\max_{j} \; (|\mathcal{N}| - j + 1) \cdot v_i + \sum \pi(m \mid g_s = i, a = 1) - \sum \pi(m \mid g_r = i, a = 1) \tag{1}$$

where the first term refers to the total value of service at the final position for goer i, the second term is the total transactions received for moving back in the queue as a requestee, and the final term is the total transactions paid for moving forward as a requester.

3.3 Simple Model of Bargaining: Complete Information

Here we provide a short synopsis of how bargaining and trade occurs in such a queue when private values of time are public knowledge and a "take-it-or-leave-it" mechanism is enforced. While the common knowledge assumption deviates from the general setting and experimental paradigm we focus on, the process provides useful qualitative insights into how the exchange within the queue may develop.

Let n goers be organized in a queue, each with some value for service v_i, as defined in Section 3. Each goer is fully aware of (1) his own private value of time, (2) his own position in the queue, (3) the positions of all other goers, and (4) the values of time held by each of the other goers (as well as the distribution over values, $F(v)$).

Since each trade includes only two goers: the goer who submits the bid and the goer who owns the position, we consider each bargaining process as a bilateral bargaining. The bidding process unfolds as follows. Knowing their own v_i and the value of their desired position in the queue (the highest available), v_{pos}, each trader bids on the best position for which $v_i > v_{pos}$. For example, the set of goers whose value is higher than the value of the position 1 goer, $\mathcal{N}_{v_i > v_{pos=1}} \equiv \{i \in \mathcal{N} : v_i > v_{pos=1}\}$, will bid on the first position. Naturally, each wishes to bid as low as possible. However, knowing this, each relevant goer will slowly increase their bid until they reach their respective v_j, resulting in a reverse Bertrand bidding dynamic. The goer with the highest value thus only needs to bid some increment, ε above the value of the second-highest goer. The winning goer, say i, settles in the first position, thus trading places with the goer initially holding the first position, say goer j (meaning goer j is relegated to the i^{th} position in the queue).

Now that the first position has been decided, the second position is up for sale (in the event that $\mathcal{N}_{v_i > v_{pos=k}} = \emptyset$, the bidding moves to position $k+1$).[1] Bidding proceeds as with the first position, with the highest valued goer among $\mathcal{N}_{v_i > v_{pos=2}}$ paying ε above the second highest relevant goer. This process continues for all n positions in the queue, leaving all positions settled with traders organized in decsending order from highest to lowest valued. If, for some k, $\mathcal{N}_{v_i > v_{pos=k}} = 1$, the singleton goer will bid $v_{pos=k} + \varepsilon$.

Under this special, common knowledge case, the gains from trade (for the initial position

[1] Note that for $k > 1$, $\mathcal{N}_{v_i > v_{pos=k}}$ is drawing from the set of bidders who have not settled in a position, meaning the size of the set from which $\mathcal{N}_{v_i > v_{pos=k}}$ is drawing is $n - k$.

holders) range from 0, in the case the goers are already ordered in descending value to $\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor}(n-2*i+1)(v_{i+1}-v_{n+1-i}+\varepsilon)$ in the case that goers are ordered in ascending value.[2] In all the cases, the reordered queue is Pareto efficient.

As an example, let a queue have six queue goers whose values for time are drawn uniformly without replacement from the set $\{2,4,6,8,10,12\}$. In the max-gains-from-trade scenario, the first and sixth position swap with the new first position holder paying $8+\varepsilon$ per position (so $5(8+\varepsilon)$) to the initial first position holder, the second and fifth position swap with $3(4+\varepsilon)$ going to the seller, and the third and forth swap with the seller gaining just ε. The gains from trade for the buyers (or requesters) can be simplified to $\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor}(n-2*i+1)(v_i-v_{i+1}-\varepsilon)$. In the above example, this corresponds to the sixth goer gaining $5(2-\varepsilon)$, the fifth goer gaining $3(2-\varepsilon)$ and the fourth goer gaining $2-\varepsilon$. In total, buyers (requesters) yield $18-9\varepsilon$ of the total gains from trade of 70, while sellers (requestees) yield $52+9\varepsilon$.

3.4 Simple Bargaining Process with Incomplete Information

Following the theoretical framework in the Section 3.3, what if the queueu-goers are only informed of their own information and the distribution of the value of time $F(v)$? Releasing the complete information assumption enables our hypotheses to be closer to the environment in our experiment.

The bidding process is similar to the previous one with complete information, but the reverse Bertrand dynamic is now replaced by a first price private value auction. The bidding starts with the highest available position and moves backward. In each bidding, the goers behave as if they are the bidders in the first price private value auction. Without loss of generality, we assume that each bidder submits one bid to one position. The goer with the

[2]Here goer indexes/identifiers are ordered in terms of where their final position should be; or, in other words, in rank order of value of time. If n is odd, then under this arbitrary indexing choice, no additional value is gained from the unpaired goer, as the holder of position $\lfloor \frac{n}{2} \rfloor+1$ does not move.

highest value wins the auction if the bid exceeds the value of service of the goer who owns the position. Then the game continues to the second highest available position and ends when all the positions have been auctioned.

Compared to the Pareto efficient queue with complete information, the reordered queue with incomplete information is less efficient. Shading occurs during the auction and thus the goers in the reordered queue may not be sorted by their descending value of service. However, adding such a trading mechanism to the queue improves efficiency compared to the initial random queue.

4 Experimental Design

4.1 Procedure

In each experimental session, 12 subjects play 12 two-minute rounds of queuing games (after 2 practice rounds). At the end of the session, two random rounds are selected to pay the subjects. In each round, the subjects are divided into two 6-player groups. We apply random matching between rounds so the subjects are re-grouped between rounds.

Figure 1 displays a sample user interface. Six subjects are randomly positioned in a virtual queue at the beginning of each session, waiting for service. In our experiment, the game starts with an inefficient queue, and we take the players' arrival time as exogenous. To control for the randomness of the position assignment, we design three random queues and circulate them between rounds. To make the efficiency improvement more apparent, in all three predetermined queue variants, the players with high service value are innitially placed near the back positions. [3] The arrangement of the three queues between rounds provides a value matrix that defines the value of each subject in each round. We control the value

[3]The value of service of the three queues are (from front to back): [2,4,6,8,10,12], [6,2,4,10,8,12], and [4,8,2,6,12,10]

matrix between experimental sessions, thus the queues are the same between sessions if they share the same round number. The subjects' service values per position are private and heterogeneous, but the subjects are informed of the distribution of the service values. When the subject's position moves forward, their total service value increases linearly. The round payoff depends on a subject's final position when the round ends, plus the subject's earnings from position exchanges when monetary transfers are allowed.

At any time during the round, the subjects can send an exchange request to any subject in front of them, and can also receive requests from the subjects behind them. Each subject can receive multiple requests simultaneously, but can only maintain one request as a requester themselves. When the subjects accept a request, the two subjects (the requester and requestee) switch positions and all the other requests related to the two subjects are automatically cancelled. The subjects can also manually cancel their own request at any time prior to acceptance/rejection.

The user interface (from the top to the bottom) displays the current round information, the queue information, the actions the subjects can make, their current sent request, the request from other subjects, their current payoff, and their exchange history. In Figure 1, we can observe that this subject is currently in round 3 with TL (take-it-leave-it auction) and no messaging treatments. The subject is at position 4 and the service value increase for a position forward is 9. The subject has not sent any request yet but has already received two requests from other subjects. From the history we can find that the subject is originally at position 5 but has traded with another subject to position 4 costing 8 lab currency.

Furthermore, as can be seen in Figure 1, the subjects can send a free text message with their requests if in a communication treatment session.

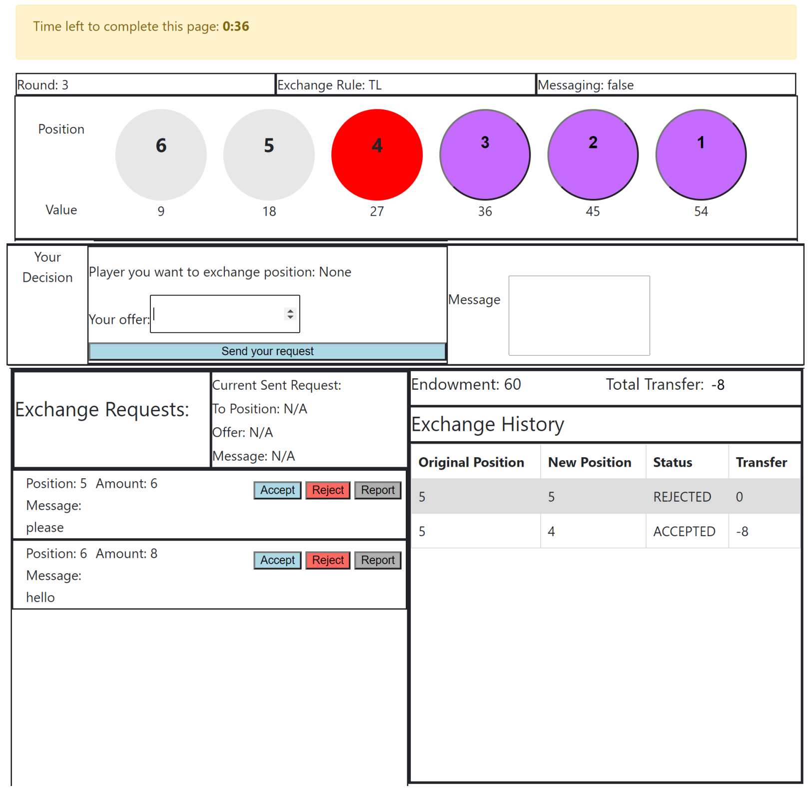


Figure 1: Sample user interface. The red dot denotes a subject's own position. Purple dots denote traders ahead of the subject in the queue. A subject's current queue messages and (round) exchange history are shown in the lower left and right sections of the screen, respectively.

4.2 Exchange rules

The paper applies a 3 $\times$ 2 full-factorial between-subjects design. In our experiment, we introduce four exchange rules.

In **Swap** treatment queues, the subjects are allowed to send requests and switch positions, but with no monetary offer. The treatment serves as the baseline of the experiment, where the motivation for switching is simply altruism. A rational subject will reject all the requests from other subjects, thus no improvement will be made to the overall efficiency.

In **TL** treatments, the subjects can make offer with their exchange requests. A *requester* (the one who makes the request) can keep only one offer at any given time, but a *requestee* (the one who receives the request) may receive multiple offers simultaneously. If an offer is accepted, a direct transfer happens between the subjects with the offered price.

In **Token** treatments, no monetary transfers are allowed between the subjects. Instead, a subject earns a token if they move backward and loses a token when they move forward. The “token” can be considered as a symbol of altruism, which could motivate egalitarianism. Note that the token can be consumed if a player moves backward and then forward, but such a motivation is rarely found in our experiment.

We also design a **Comm**unication treatment where the subjects are allowed to attach a free-text message to their requests. The requestees can observe the message together with the requests. Note that the Comm treatment is not a parallel treatment with the other three rules, but is combined with them, which gives us 6 combinations of treatments.

4.3 Sessions

Exchange rule	Communication	# sessions (# subjects)
Swap	Yes	2(24)
Swap	No	2(24)
TL	Yes	2(24)
TL	No	2(24)
Token	Yes	2(24)
Token	No	2(24)

Table 1: Session information.

The data were collected during 12 online sessions conducted between October 2021 and March 2022 with 144 subjects. Subjects were recruited through the Experimental Economics Laboratory at Universidad del Pacifico on Orsee (Greiner (2015)). The session information is shown in Table 1. The experiment is developed on oTree (Chen et al. (2016)). Each session lasted about an hour with an average payment of 16 PEN per subject (approximately $4 USD).

4.4 Hypotheses

In this section, we develop hypotheses to guide our experimental study. As alluded to above, a setting with risk neutral agents would predict that queues in the Swap treatment will yield no adjustments from FCFS and thus have the same level of efficiency, giving us a baseline hypothesis:

Hypothesis 0. *The gain in efficiency for Swap - No Communication queues is expected to be zero, with no trades occurring, yielding equivalent outcomes to FCFS.*

Compared to the baseline environment where players can only exchange without extra incentives, adding the take-it-or-leave-it auction to the queuing problem obviously can improve market efficiency. As we have discussed in Section 3, with a complete information environment, the queue is efficient. With a incomplete information in our experiment, we expect the queue efficiency to be between the baseline and the efficient queue.

For the subjects who value the symbol of altruism, they are also more likely to switch backward in the Token treatment.

Hypothesis 1. *Compared to the queue efficiency in the Swap treatment, the efficiency is higher (a) in the TL treatment and (b) in the Token treatment.*

The communication in our experiment is considered as cheap talk and is not self-committing. Compared to the Token treatment where people receive tokens for their behavior, the requesters can disclose their service value, but deception is also possible. As a baseline, we hypothesize that the messages alone cannot improve queue efficiency.

Hypothesis 2. *Cheap talk does not improve queue efficiency.*

When considering the exchange in the TL treatment, we are interested in how the surplus is allocated between the requesters and the requestees. Hypotheses 1 is developed based on the fact that the requestees initially own the front positions. These positions are considered as the subjects' initial endowment, thus bringing extra payoff to them if the positions are traded. Furthermore, the requestees can receive multiple offers simultaneously but each requester can send at most one offer at the same time, thus brings more bargaining power to the requestees.

Hypothesis 3. *When monetary trade is allowed, the requestees earn a higher surplus on average than the requester.*

5 Result

5.1 Data Overview

We have two sessions for each combination of treatments, thus generating 48 groups (2 groups in each round with 24 rounds in total) of data per treatment. Table 2 shows the summary of main results between treatments. To compare the efficiency improvement, we normalize the efficiency (ratio between group total payoff and group's maximum possible payoff) to $[0, 1]$, where 0 represents the efficiency of the initial random queue, and 1 refers to the maximum possible efficient queue. The requester/requestee share is calculated by the difference between the subjects' service value and the offered price per position, which clearly displays how much the subjects earn from the exchanges. In Swap and Token treatments, the monetary transfer is not allowed, thus the share is only calculated by the subjects' service value per position. Since the requestees will move backward with no monetary compensation, their average share is negative in the Swap and the Token treatments.

The table reveals a clear distinction between treatments. Although the subjects accept more requests in the Token treatments than in the Swap treatments, the efficiency and

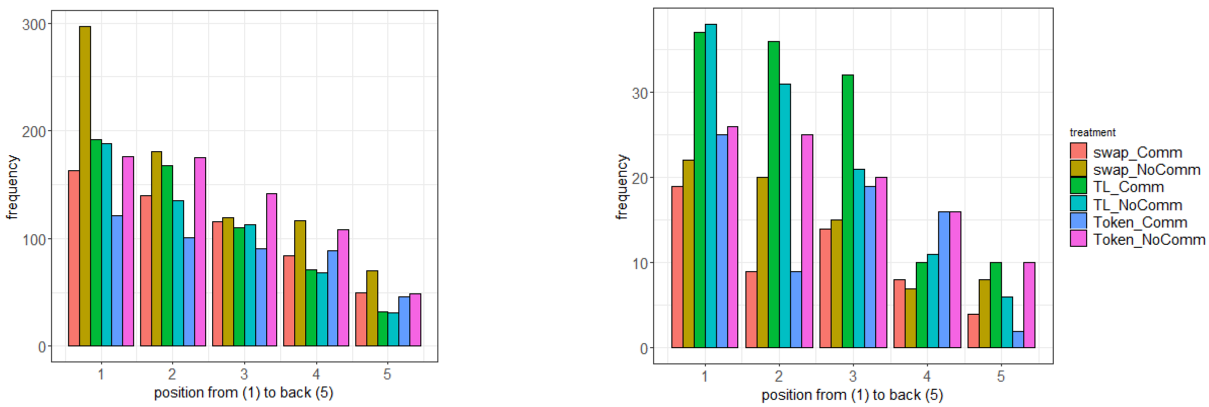
	swap_NoComm	swap_Comm	TL_NoComm	TL_Comm	Token_NoComm	Token_Comm
accept rate	0.09	0.10	0.20	0.22	0.15	0.16
efficiency	0.05	0.08	0.33	0.37	0.10	0.05
#groups	48	48	48	48	48	48
#asks/groups	16.33	11.54	11.15	11.96	13.54	9.33
#messages/groups	-	7.85	-	1.98	-	5.56
requester gain	7.91	7.99	0.49	1.20	7.26	7.26
requestee gain	-5.54	-5.43	1.19	0.16	-4.77	-4.91
requester gain accepted	6.69	8.15	-0.53	0.16	6.62	6.14
requestee gain accepted	-5.22	-5.48	4.94	3.79	-4.44	-4.88

Table 2: Summary data.

allocation between requesters and requestees do not change much, which shows that the additional token coins trigger more altruism and egalitarianism but do not improve the market efficiency. It is also not surprising that the efficiency improvement in the Swap and the Token treatments is low, since the acceptance rate is low and rational players have no incentive to exchange positions. Adding communication does not affect the efficiency significantly, as the free-text messaging is simply cheap talk in our experiment.

On the other hand, subjects accept more and efficiency improves in the TL treatments when the auction is allowed, compared to the Swap and the Token treatments. In addition, from the TL treatment we can find evidence that the requestees earn higher in the accepted exchanges, which indicates that the requestees have a high market power. The result is reasonable in this experiment, as the initial position can be considered as the players' endowment and the requestees can receive multiple offers from the requesters simultaneously.

Figures 2 and 3 show details about the frequency of requests and the average offer per position. The front positions are more attractive and receives more requests. But the subjects do not bid higher for front positions. They bid higher for the back positions instead when the total transfer is small.



(a) All requests

(b) Accepted requests

Figure 2: The frequency of requests by requestees' position.

5.2 Exchanges in TL treatments.

The exchange requests can be very diverse since there is no limit on the requests. As such, we are interested in the patterns among the accepted requests. If the requesters have a higher service value than the requestees, the overall efficiency increases when the request is accepted, and both subjects earn money when the offer price is between the service values of the two subjects involved. Do the subjects always earn money in the exchanges? At the outset, neither player loses in 67.8% of the exchanges, the requester loses money in 27.0% of the exchanges, and the requestees lose money in 3.4% of the exchanges. Both players lose in 1.7% of the exchanges. In this section, we investigate the first two most frequently played scenarios.

We first look at the case when both players do not lose money. In the previous section, we have find evidence that the requestees tend to have a higher share in the TL treatment. Figure 4 reviews the share between the requesters and the requestees by treatments. To make the exchanges at different scales comparable, we normalize the share to $[0, 1]$, where the subject earn nothing at 0, and earns all the surplus from the exchange at 1. In both TL treatments with and without communication, we find that the requestees tend to have a large share of the surplus at around 1, where the requester tends to earn little. The figure

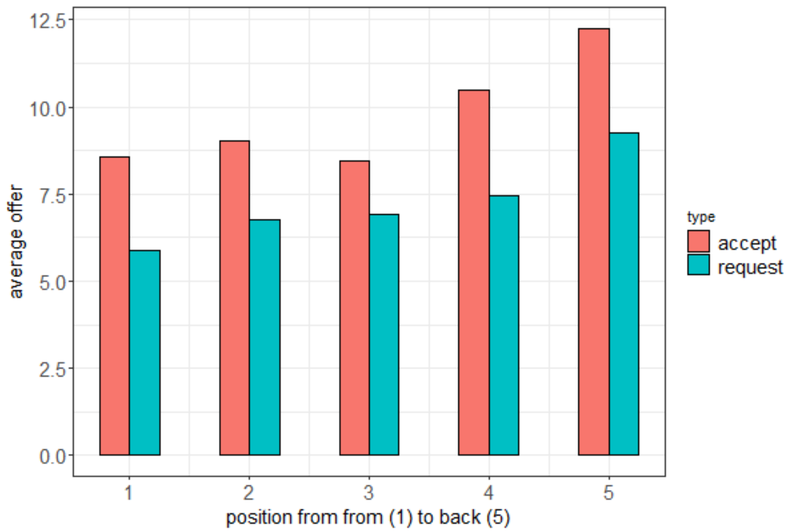


Figure 3: Average offer by requestees' position.

confirms our inference in the previous section.

When the requester submit an offer which could make them lose money, the experimental program sends them a warning message so the loss is not resulted from careless or miscalculation. What makes the requester insist on losing money for an exchange? "Joy of winning" is a possible explanation. The motivation of reselling the position at a higher price could be another reason. However, from Figures 5 and 6, we learn that the these requesters tend to lose money at the end of the round, and these types of requests are in fact sent by a few subjects (7 subjects contribute 62% of the total such exchanges, 39 out of 63 requests). By tracking the following exchanges of these subjects, we find that these subjects rarely earn money later in the round. In fact, the majority of these subjects do not accept any offer after the exchange. For the subjects with follow-up exchanges (around 30%, 12 out of 39 requests), they either continue to move to a higher position with negative exchange value, or they resell the position at a lower price. In this case, "Joy of winning" seems to be a reasonable explanation for the irrational trade.

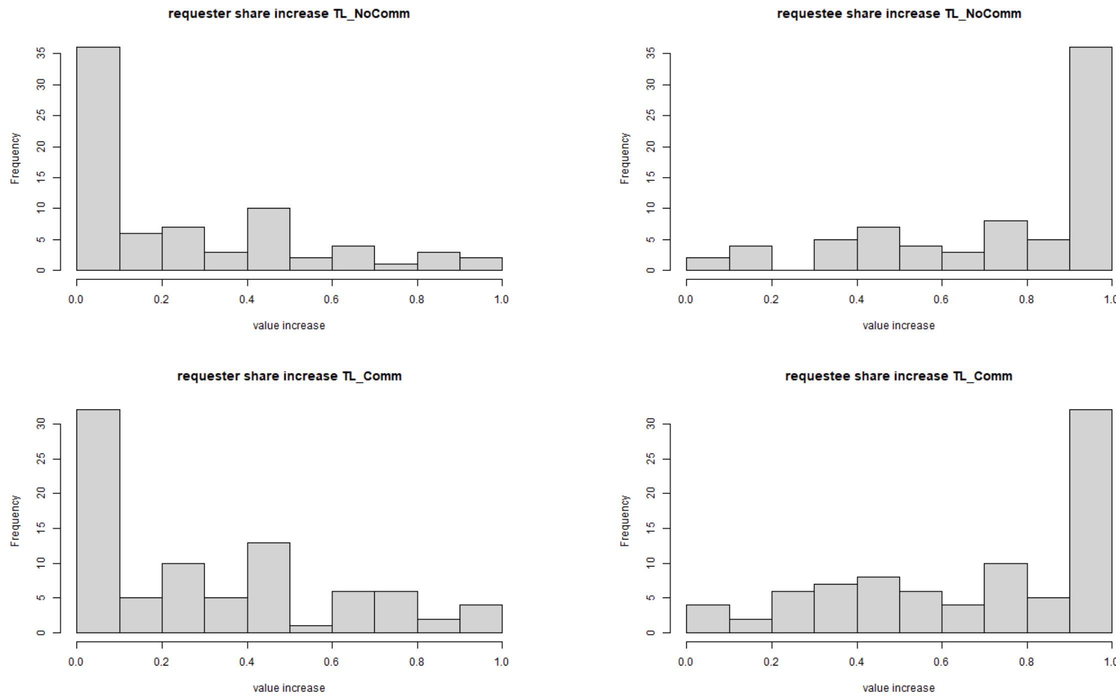


Figure 4: value increase per position from the accepted exchange.

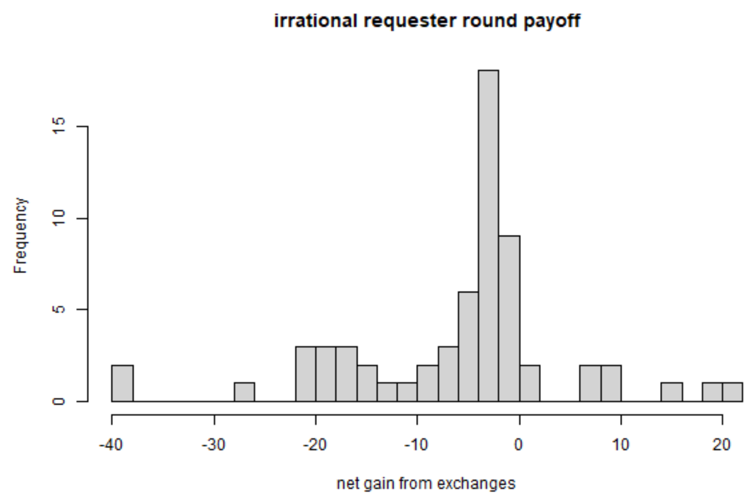


Figure 5: Round payoff for the irrational requester.

5.3 Treatment Effects

We support many of the tendencies shown descriptively above via regression analysis as well. Table 3 reports treatment effects on the action (ask and message) frequency at the per

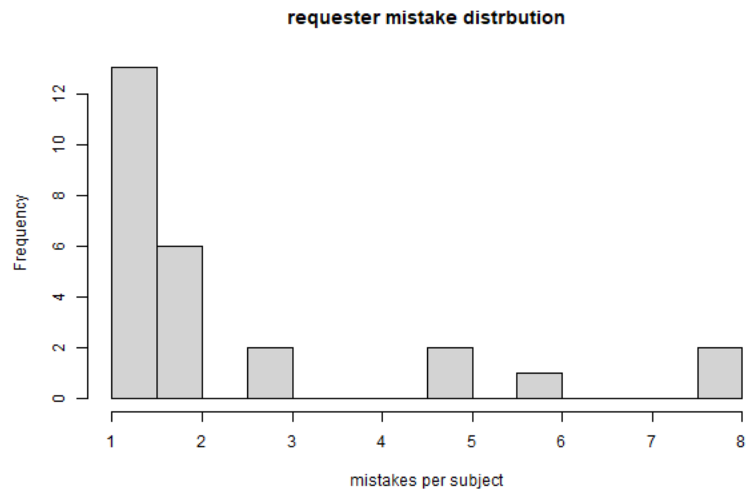


Figure 6: The distribution of requesters making mistakes.

player level and two major measures of queue health (acceptance rate and efficiency). We see matching signs across dependent variables for the token and take-it-or-leave-it main effects; each increases acceptance rate and lowers both asks and messages per player. Both are indicative of players taking more time to interact with and respond to the more information-rich, trade-incentivizing queues. Interestingly, magnitudes are consistently and substantially higher for TL queues across the board. As such, noticeable improvements in efficiency are found in TL queues, while Token queues reveal small, yet insignificant gains. Chat main and interaction effects support the tendencies found previously; namely, a near one-to-one trade off between messages and asks is found in token queues, while TL queues see no real change.

With Table 3, we can finally answer hypotheses 1 and 2 with some additional findings.

Result 1. *(a) The Swap treatment provides a low, but significant non-zero efficiency improvement. (b) Compared to that in the Swap treatment, the queue efficiency in the TL treatment is higher but the same result is not statistically significant for the Token treatment. Although the acceptance rate is higher in both treatment.*
(b) Adding an additional treatment significantly reduce the number of asks and the frequency of messaging.

	(1)	(2)	(3)	(4)
	Acceptance Rate	Asks Per Player	Messages Per Player	Efficiency
Token	0.067***	-0.465**	-0.382***	0.050
	(2.74)	(-2.17)	(-3.92)	(1.35)
TL	0.103***	-0.865***	-0.979***	0.270***
	(4.19)	(-4.03)	(-10.05)	(7.33)
Chat	-0.004	-0.799***		0.025
	(-0.16)	(-3.72)		(0.67)
Token:Chat	-0.002	0.097		-0.074
	(-0.06)	(0.32)		(-1.42)
TL:Chat	0.029	0.934***		0.021
	(0.83)	(3.08)		(0.41)
Round	0.001	0.058***	0.029**	0.006*
	(0.51)	(3.24)	(2.47)	(1.96)
Intercept	0.093***	2.228***	1.067***	-0.000
	(3.77)	(10.36)	(8.91)	(-0.00)

t statistics in parentheses. Baseline condition: swap method, no communication.

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

Table 3: Actions per player and Measures of Queue Success.

Result 2. *Cheap talk does not improve queue efficiency in our experiment.*

Much like with auction and queue performance, gains from trade reveal similar trends between the two more sophisticated mechanism queues relative to swap queues. Table 4 reports the findings, with the first two columns including (expected) gains for all requests and the last two using only requests which the requestees accept. The intercepts reveal the average per position value of time for requesters (1 and 3) and requestees (2 and 4). As the queues are predisposed to having higher time costs at worse queue positions, moving from (1) to (3) shows requesters are successful more often when requesting from positions further from the rear of the queue.[4] We see that across all requests, TL requesters bid just slightly under their value, on average nearly truthfully bidding. Successful requesters appear to bid nearly 20% above their time value on average, though this is a result of having lower values, not higher bids. A three unit premium appears in TL sessions for requestees to accept a bid over the average. This leads to requestees receiving about twice their value per position to move back in the queue. Token queue estimates reveal matching signs to TL queues, though an order of magnitude or more less in size, and mildly significant at best.

Inclusion of in-queue messaging seems to offer an interesting tool for requesters to remove some of the requestees' leverage in TL queues. While requester gain is essentially unchanged, requestees gain are significantly reduced in both samples. Most of the expected gain seen across all requests for the requestees is removed. This persists in accepted requests, though to an even larger extent, with a third of the net gain from moving back disappearing for requestees. A natural line of intuition for this would be that requesters are effectively able to reduce the requestees' premium, whether through value revelation or cheap talk.

From Table 4, we can answer Hypothesis 3 with the following.

Result 3. *(a) On average, the requestees take the majority of the surplus in the exchange under the TL treatments.*
(b) Though has an little impact on the queue efficiency, messaging reduces the requestees' premium in the TL treatments.

5.4 Message Categorization

To better address the experimental findings when messaging is allowed, we perform a text analysis regimen on the full set of in-queue messages. In Table 5, each column refers to a

[4] Unsurprisingly, the intercepts for requestees are essentially the same.

	(1)	(2)	(3)	(4)
	Requester Gain	Requestee Gain	Accepted Requester Gain	Accepted Requestee Gain
Token	-0.540*	0.685**	-0.268	0.584
	(-1.95)	(2.43)	(-0.36)	(0.92)
TL	-7.498***	6.915***	-7.864***	10.500***
	(-27.04)	(24.50)	(-10.63)	(16.81)
Chat	0.097	0.125	1.623**	-0.156
	(0.35)	(0.44)	(2.07)	(-0.23)
Token:Chat	-0.128	-0.244	-2.073*	-0.042
	(-0.33)	(-0.61)	(-1.94)	(-0.05)
TL:Chat	0.614	-1.153***	-0.010	-1.675*
	(1.57)	(-2.89)	(-0.01)	(-1.90)
Round	-0.014	-0.021	0.016	0.019
	(-0.59)	(-0.87)	(0.27)	(0.37)
Intercept	8.007***	-5.331***	6.593***	-5.202***
	(28.81)	(-18.84)	(8.34)	(-7.80)

t statistics in parentheses. Baseline condition: swap method, no communication.

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

Table 4: Gains from Trade at the per trade per queue level.

specific information category that the messages contain. Note that a message can belong to multiple categories, or no category at all. The numbers show the fraction of messages that belongs to each category. We select seven typical message categories as follows.

- “With offer”: the messages includes an explicit mention of the offer in numeric term;
- “Explicit request”: the messages explicitly states a request;

- “Mention surplus”: the messages mentions a gain for the message recipient associated to accepting the offer;
- “Mention own value”: the messages mentions own potential gain of the transaction;
- “Persuasive”: the message appears to be persuasive in a sales people or effective emotion type of way;
- “Deception”: the message is considered to attempt deception or to reflect a misunderstanding;
- “Emotion”: if the messages contains emotional persuasion.

	swap_Comm	Token_Comm	TL_Comm
#messages	377	267	95
With offer	0.65	0.65	0.87
Explicit request	0.42	0.57	0.09
Mention surplus	0.21	0.03	0.18
Mention own value	0.16	0.06	0.00
Persuasive	0.38	0.22	0.23
Deception	0.06	0.30	0.01
Emotion	0.14	0.08	0.01

Table 5: Message type summary.

From Table 5, the majority of the messages successfully contains an explicit request or an offer, which clearly shows the requester’s intention. However, the requester often fail at sending informative and convincing requests. In the Swap and Token treatments, the requestee is guaranteed to lose value in the current exchange. We found that the requester tend to be more persuasive, emotional, and deceptive. On the contrary, the requester tends to discuss the surplus in the TL treatment, in which the actual value transfer involves.

6 Conclusions

While first-come-first-serve queues are often accepted as being a fair, though inefficient, option among the general public, and centralized mechanism alternatives have often been presented to improve efficiency, more externally applicable decentralized mechanisms have been proposed and tested scarcely. We test three simple decentralized mechanisms (Swap, Take-it-or-Leave-it, and Token) in settings with and without the ability to communicate directly with others. In our laboratory experiments, monetary exchange won out in terms of gains in efficiency relative to FCFS, backing an auction-based, as opposed to egalitarian route to improving queues. In addition, the requestee tends to take the majority of the surplus and some requesters systematically lose money during the exchanges.

A few relevant implications from our experimental analysis are as follows. First, though the queuing system in our experiment provides the requestee with more bargaining power, a few requesters are willing to lose money for the “joy of winning”, which indicates that the front position itself seems to affect the players utility beyond the service value it can provide. Second, the efficiency fails to increase in the Token treatments. The token in our queuing system serves as a representation of altruism with heterogeneous values. The failure of the token system also indicates the failure of building a temporary altruistic community in the laboratory. Future research on topic could also incorporate more dimensions (ie decentralized in two dim, position and quality (eg sports tickets))

References

Bigoni, M., Camera, G., and Casari, M. (2020). Cooperation among strangers with and without a monetary system. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing.

Chen, D. L., Schonger, M., and Wickens, C. (2016). otreean open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97.

Dewan, S. and Mendelson, H. (1990). User delay costs and internal pricing for a service facility. *Management Science*, 36(12):1502–1517.

Edelson, N. M. and Hilderbrand, D. K. (1975). Congestion tolls for poisson queuing processes. *Econometrica: Journal of the Econometric Society*, pages 81–92.

El Haji, A. and Onderstal, S. (2019). Trading places: An experimental comparison of reallocation mechanisms for priority queuing. *Journal of Economics & Management Strategy*, 28(4):670–686.

Farrell, J. and Rabin, M. (1996). Cheap talk. *Journal of Economic perspectives*, 10(3):103–118.

Giebelhausen, M. D., Robinson, S. G., and Cronin, J. J. (2011). Worth waiting for: increasing satisfaction by making consumers wait. *Journal of the Academy of Marketing Science*, 39(6):889–905.

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

Hassin, R. (1995). Decentralized regulation of a queue. *Management Science*, 41(1):163–173.

Kayı, Ç. and Ramaekers, E. (2010). Characterizations of pareto-efficient, fair, and strategy-proof allocation rules in queueing problems. *Games and Economic Behavior*, 68(1):220–232.

Kleinrock, L. (1967). Optimum bribing for queue position. *Operations Research*, 15(2):304–318.

Koo, M. and Fishbach, A. (2010). A silver lining of standing in line: Queuing increases value of products. *Journal of Marketing Research*, 47(4):713–724.

Kremer, M. and Debo, L. G. (2012). Herding in a queue: A laboratory experiment. *Chicago Booth Research Paper*, (12-28).

Kumar, P., Kalwani, M. U., and Dada, M. (1997). The impact of waiting time guarantees on customers' waiting experiences. *Marketing science*, 16(4):295–314.

Mendelson, H. and Whang, S. (1990). Optimal incentive-compatible priority pricing for the m/m/1 queue. *Operations research*, 38(5):870–883.

Naor, P. (1969). The regulation of queue size by levying tolls. *Econometrica: journal of the Econometric Society*, pages 15–24.

Pruyn, A. and Smidts, A. (1998). Effects of waiting on the satisfaction with the service: Beyond objective time measures1. *International journal of research in marketing*, 15(4):321–334.

Rapoport, A., Stein, W. E., Parco, J. E., and Seale, D. A. (2004). Equilibrium play in single-server queues with endogenously determined arrival times. *Journal of Economic Behavior & Organization*, 55(1):67–91.

Seale, D. A., Parco, J. E., Stein, W. E., and Rapoport, A. (2005). Joining a queue or staying out: Effects of information structure and service time on arrival and staying out decisions. *Experimental Economics*, 8(2):117–144.

Stein, W. E., Rapoport, A., Seale, D. A., Zhang, H., and Zwick, R. (2007). Batch queues with choice of arrivals: Equilibrium analysis and experimental study. *Games and Economic Behavior*, 59(2):345–363.

Yang, L., Debo, L., and Gupta, V. (2016). Trading time in a congested environment. *Management Science*, 63(7):2377–2395.

Zhou, R. and Soman, D. (2008). Consumers' waiting in queues: The role of first-order and second-order justice. *Psychology & Marketing*, 25(3):262–279.