Efficiency in Queuing Under Decentralized Mechanisms*

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

May 18, 2023

Abstract

We study, theoretically and in the laboratory, three simple decentralized mechanisms to reallocate positions in a queuing problem. In our environment, players have heterogeneous values for time and arrive in random order to a queue before service starts. While waiting for service, they can switch positions, with different rules depending on the mechanism in place. We mainly focus on three institutions: voluntary swapping upon request, take-it-or-leave-it (TIOLI) monetary offers, and a non-fungible reputation point system ("social token"). Compared to the initial order of the lines, we find modest efficiency gains when swapping and the social token are in place. In contrast, we find a sizable efficiency improvement in queues with monetary transfers. Although TIOLI improves efficiency, this gain comes at the cost of higher inequality. We also study these mechanisms with and without unilateral communication from the switch requester. This form of cheap talk has no impact on the main studied outcomes.

JEL Classification: C92, D47, D82.

Keywords: Queuing, Decentralized mechanisms, TIOLI, Auction, Indirect Reciprocity.

*We thank Alex Nguyen for their excellent programming work at the UCSC LEEPS Laboratory. We also thank Grecia Barandiarn, Paula Armas, and Martin Sanchez for excellent research assistance. We are grateful for feedback from the audience at the UCSC Experimental Economics Workshop and at the DUFE-IAER seminar.

†Economics Departments, University of California Santa Cruz; klopezva@ucsc.edu (corresponding author);

‡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 congestion or temporary excess demand, queuing is one of the most commonly used institutions for firms and governments to allocate scarce resources. This mechanism is perceived as intuitive and is widely adopted in housing allocation, access to sports facilities, concert ticketing, and many other industries and contexts. Understanding the properties of queuing institutions or mechanisms to improve their efficiency has occupied researchers in economics, computer science, and mathematics. Yet, mechanism design for queuing problems still faces important open challenges.

Usually, customers are served on a first-come-first-serve (FCFS) basis. FCFS is perceived as fair, as participants seemingly arrive at their own time, but it is often socially suboptimal.[1] Regularly, the order of the line does not perfectly correlate with the ordering of the value of time, sometimes generating large inefficiencies. A big deal of attention has been given to centralized mechanisms since these can theoretically solve the inefficiency problem (Kleinrock, 1967; Yang et al., 2016, e.g.). However, they are not perceived as simple nor fair by typical humans relative to the standard FCFS mechanism (Zhou and Soman, 2008). Decentralized mechanisms, defined as institutions where queue participants deal with each other without the assistance of a server or other authority, have received much less attention.

In this paper, we study several decentralized trading mechanisms theoretically and in the laboratory. These mechanisms are relatively simple or intuitive and, therefore, could be learned and adopted by humans with relative ease. In this paper, as the first step to assess the likelihood that any of these mechanisms become a widely adopted institution, we aim to understand whether any of them can substantially improve the efficiency of queues relative to FCFS. We acknowledge, however, that studying their introduction and voluntary adoption goes beyond the scope of this paper.

In our studied environment, individuals have heterogeneous values of their time (or,

[1]FCFS principle is also know as *first in first out* FIFO.

equivalently, different costs of waiting) and arrive to be served in random order. Their time value is private, but the distribution of these values is common knowledge. Since values and initial positions are random, using FCFS is most likely inefficient. We do not model the decision to enter the line so that we can concentrate on the reordering behavior.[2]

We study three mechanisms that allow reordering upon bilateral agreement. First, we study a mechanism where two players in the queue can only swap their positions upon the request of a player to another player that is ahead in the line and the acceptance by the latter. We call this institution *Swap*. We then consider a conventional take-it-or-leave-it mechanism (henceforth, *TIOLI*) where any player at the back of the line (second or a worse position) can directly make a monetary offer with their exchange request to someone ahead of them and this person can accept or decline the offer.

As a third mechanism, we consider an option similar to the *Swap* institution, except now, players earn points for their altruistic behavior (when they let another take their place) and lose points when they request and succeed in moving forward. We call this treatment "Social Token", or *SocTok*. The intuition behind this institution is to have an observable measure or signal of individuals' pro-social and trusting attitudes based on their previous actual behavior. If people seek a more cooperative outcome where high-time-value people sincerely request a swap and get it, then SocTok may provide a means for such a reputation-driven exchange. The notion of the *SocTok* treatment is partly analogous to the intrinsically worth-less tokens in Bigoni et al. (2020), where players cooperate in repeated prisoner's dilemma games with non-monetary tokens.

We present a simple model of decentralized position trading that follows the essential elements of the environment and mechanisms described above. The model allows us to provide relevant theoretical insights and predictions mainly under the assumption of selfish preferences and complete information. It predicts that efficiency gains of the Swap and SocTok environments will be low if any with respect to FCFS; while TIOLI will exhibit

[2] We acknowledge that a full picture of the performance of the mechanisms would involve studying the impact of the mechanisms on balking.

much larger efficiency gains although there is a multiplicity of paths in which efficiency is reached and therefore it's effects on inequality are ambiguous. We also present and discuss, although less formally, conjectures on settings with repeated play and non-selfish agents.

Additionally, in the experiment we study the impact of unilateral communication in this environment. Cheap talk has been shown as an environment feature that can improve efficiency and help reach equilibrium in some contexts (Farrell and Rabin, 1996, e.g.). It is also commonly observed in our daily lives when people try to communicate with others and move forward in lines, thus worth our investigation. In our environment, cheap talk could appeal to the players' altruistic or efficiency motivations, although the communication itself is not self-committing (Andreoni and Rao, 2011). We study each mechanisms with and without verbal communication possibilities. We allow for a "cheap talk" by letting players attach a message with their requests under Swap, TIOLI, and Social Token treatments. We then test whether this form of communication impacts efficiency.

The experimental results are interesting. Acceptance rates for all three mechanisms are statistically positive. TIOLI exhibits the largest acceptance rate, followed by SocTok; which in turn has a larger rate than Swapping. This ordering arguably suggests that material benefit is preferred to reputational benefit, which in turn, is preferred to no material benefit at all. In terms of efficiency, all treatments increase the efficiency of the line. However, *Swap* and *SocTok* only do so to a small degree (they generate an increase of 6-7 percentage points relative to the initial efficiency level). The largest increase in efficiency can be attributed to the *TIOLI* mechanism, which increases by 34 percentage points. Furthermore, although the lines are more efficient in TIOLI, most of the "gains from trade" go to the players who own the front positions. This increases inequality. Whether unilateral communication is allowed or not has little to no effect on the primary outcomes we study. We also find that, in a few exchanges, some requesters propose unprofitable offers. We argue there is mixed evidence of speculation intentions or intrinsic value for (getting) the front positions (joy-of-winning).

This paper lies in the intersection of the research bodies on queuing and simple mechanisms for bilateral trade. For the last 60 years, most of the research on queuing within economics has concentrated on understanding the efficiency implications of the queue system imposed by a designer (Kleinrock, 1967) or the arrival process determined by the server or customer (Naor, 1969).[3]

Our study is closer to the server/customer literature in terms of research method, as the majority of the experimental literature is on the arrival behavior in the queue, and closer to the designer literature in terms of question and motivation, as our focus is on studying the improvement over FCFS efficiency due to the adoption of different institutions.

Since the introduction of arrival choice on the part of the customer (Naor, 1969), a swath of studies has considered various types of arrival processes. Notably, Rapoport et al. (2004) and Seale et al. (2005) present lab studies of queues with a single-server and endogenous arrival without balking. They find that arrival times cluster around salient timestamps and document that behavior closely matches the mixed-strategy equilibria predicted by the model at the aggregate level.

The closest to our study in the literature on arrival decisions is Oberholzer-Gee (2006), which studies cutting in queues in a field experiment. They find allowing transfers does imply a reordering with respect to FCFS. However, many people forego the monetary reward offered for the cut by another person, indicating that many prefer to stick to the FCFS principle. Our paper can measure efficiency gains precisely as we induce heterogeneous values for each time unit. In that sense, we can provide a clearer picture of efficiency gains associated with simple trade institutions.[4]

Furthermore, our study is unique among these experiments, as we study behavior in the queue after any arrival process (or balking) has occurred and before service takes place. To our knowledge, ours is the first laboratory experiment to study queuing behavior in such a

[3] Allon and Kremer (2018) provide a survey of the literature in Chapter 9 of the handbook edited by Katok et al. (2018).

[4] There is a literature of experimental studies on the psychology of waiting and the behavioral impacts of adjusting the waiting mechanics and queue interfaces.See section 9.2 of Katok et al. 2018, written by Allon and Kremer 2018 for a detailed discussion.

setting.

In the theory literature, Kleinrock (1967) showed efficiency in the queue could be achieved if the designer allows for customer-to-server payments depending on their positions– and several mechanisms have been theoretically proposed since. Many such mechanisms consider trade for priority via either pricing (Mendelson and Whang, 1990; Afeche and Pavlin, 2016, e.g.,) or auctions (e.g., Afeche and Mendelson (2004)).

Closer to our paper, Rosenblum (1992) and Yang et al. (2016) consider time-trading mechanisms whereby agents develop bidding processes dependent on their wait times. This paper joins El Haji and Onderstal (2019), which compares the efficiency between a server-initiated auction and customer-initiated trade, as one of the few articles that experimentally explore efficiency adjustments and behavior exhibited in a queue when trade (or swapping) is allowed amongst customers.[5] Our paper, considers more decentralized mechanisms and explores the role of communication in different mechanisms with transfers of material or reputationally transfers.

Finally, our study of the TIOLI mechanism relates to the literature on bilateral bargaining. Specifically, our findings add to the list of bilateral bargaining results under incomplete information that started by (Chatterjee and Samuelson, 1983). Again, our paper joins El Haji and Onderstal (2019) as the only papers experimentally studying bilateral trade with incomplete information in a queue. However, we focus on the TIOLI mechanism as opposed to a double auction.

The rest of the paper is organized as follows. In Section 2, we build a simple model of queuing to guide our experimental hypotheses. 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.

[5]Shunko et al. (2018) provides the only analogous experimental test on the server side of the literature.

2 A Simple Model

Consider a set of queue-goers, $\mathcal{N} = \{1, 2, 3..., n\}$, with the capacity 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. Formally, $\mathcal{Q}$ is a vector of positions where the i'th entry is the position of the i'th player. We write j_i to denote i's initial position in the line. We assume the queue goers' initial position is randomly determined before the game and that there are no empty positions in the line.[6]

For simplicity, unless stated otherwise, we assume that everyone values the service equally $\mu_i = \mu$, although this assumption is not essential. Additionally, the service time is the same for everyone and is normalized to one unit of time. Let $\mathbf{J}$ denote the vector of positions describing the current state of the queue at some point in time before service, where the ith entry, j_i, denotes the i's position in the queue in that state.

Incentives for changing positions come from people's heterogeneous value of time (or the opportunity cost of waiting one unit of time), which makes being at the back of the line undesirable, ceteris paribus. In particular, each goer i has a private value for a unit of time, v_i, drawn from a distribution $F(v)$. In sum, for player i, the utility of being in position j_i is given by:

$$u_i(j_i|v_i) = \mu_i - j_i v_i \tag{1}$$

That is, the utility associated with a position is the value of the service minus the cost of the waiting time until the end of the service. Note that the value of moving one place ahead, $u_i(j|v_i) - u_i(j-1|v_i)$, is simply v_i. In sum, Nature distributes time values and initial order in the line, then players interact with one another according to the specified rules.

It is also helpful to define and characterize the *cooperative outcome*. Note that, before being informed about their value of time, players are symmetric. We, therefore, define

[6]That is, the length of the queue equals the number of queue-goers ($|\mathcal{Q}| = |\mathcal{N}|$).

cooperative outcome as the state of the line that maximizes the ex-ante utility of any player: $\mathbb{E}[\mu - j_i v_i]$. It is immediate to see that this expectation is maximized the positions are assigned in reverse order with respect to time values. Formally, if $v_{(k)}$ is the kth statistic of time values, we want to give this value the $(n+1-k)$th position in the line. In this way, the maximum value receives the first position, the minimum value receives the last position, and everyone in the middle receives their position according to their value, in reverse order. We denote this state of the line with $\mathbf{J}^*$. Notably, the cooperative outcome in this context is also the most egalitarian state (provided that $\mu_i = \mu$ for all i). To see this, notice that $\mathbf{J}^*$ minimizes the distance among all waiting costs $j_i v_i$.

Finally, we will later state some results for the case of non-selfish preferences. Therefore, we need to define preferences over the state of the line, $\mathbf{J}$, and the realized vector of values, $\mathbf{v}$. We assume the shape of possible social preferences is given by $f(\cdot)$ so the decision utility takes the form:

$$U_i(\mathbf{J}|\mathbf{v}) = f\left((\mu - j_k v_k)_{k\in\mathcal{N}}\right) \tag{2}$$

where j_k is the kth entry of line state $\mathbf{J}$ and $f(\cdot)$ is a weakly increasing function with respect all entries. The case of selfish preferences is when $f(\cdot)$ is constant with respect to any entry $k \neq i$. Player i's preferences are said to be *altruistic* if $f(\cdot)$ is strictly increasing for at least one $k \neq i$. We say i exhibits *impartial* preferences, in the sense that player i treats any two players equally, if $f(\cdot)$ is independent of the identity of players. That is, if we swap the identities of players k and k' inside set $\mathcal{N}$ so that their entries in $f(\cdot)$ are also swapped, the value of $U_i(\cdot)$ is unaltered. Note that a large class of social preferences satisfy *impartiality.* This class includes perfectly egalitarian (Rawlsian) preferences ($f = \min_i\{\mu - j_i v_i\}$) as well as perfectly efficiency-oriented utilitarian ones (e.g., $f = \sum_i^n [\mu - j_i v_i]$).

We study three decentralized position-trading mechanisms in this environment: swapping, take-it-or-leave-it offers, and the social tokens described above. The rest of the section provides a theoretical benchmark under the assumption of complete information.

2.1 Swapping

The *swapping* mechanism allows any player i to request another player i', with $j_i > j_{i'}$, to swap positions. Player i' can, in turn, accept or reject the proposal. If and only if it is accepted, the positions are swapped as solicited by i, and payoff changes are $+(j_i - j_{i'})v_i > 0$ for player i and $-(j_i - j_{i'})v_{i'} < 0$ for player i'.

Under the assumptions of selfish preferences, complete information, and if the game is played for a finite number of periods, it is not individually rational to accept any proposed position trade. Therefore, the only Nash equilibrium is to stick to the default method, first-come-first-serve.

What would happen if preferences were not selfish? The equilibrium might change dramatically. If preferences are altruistic and impartial, the cooperative, socially-optimal outcome can be reached. This outcome, in this context, is also the most egalitarian. The rational play under those preferences states that it is any player i''s optimal behavior to accept a swapping request from player i if and only if $v_i > v_{i'}$. The intuition is that, under altruism and impartiality, the position swap will always generate opposing forces on the utility. Furthermore, as long as it is the higher-value person who is moving forward, there will be a net reduction in waiting for costs across players. Additionally, this is always true regardless of who requests the swap.

With selfish preferences, the cooperative outcome can still be achieved if the game is played infinite periods and the positions players arrive at are random and independent across periods. The cooperative play here would be: at every period, i requests a swap to move forward if her value of time is higher than the person being requested, i'; and i' accepts any request if and only if the requester has a higher value for time than herself. The cooperative outcome can be reached and stay in equilibrium. For example, every player could deploy a grim strategy if a noncooperative action takes place (an efficiency-improving swap proposal is denied or if a player who should be moving forward does not request to swap), then everyone stops asking or accepting swap requests forever. The intuition is that, since players do not know their positions in the future, they all only form expected discounted payoff, which is higher under the cooperative play than under the perpetual FCFS default process. The following proposition summarizes these theoretical insights.

Proposition 1. *In the environment described above, under complete information:*

(a) If preferences are selfish and the queue is repeated for a finite number of periods, the only Nash equilibrium is not to accept any swapping request.

(b) If preferences are altruistic and impartial, cooperative play is a Nash equilibrium regardless of the number of periods the queue is played.

(c) If the queue is repeated infinitely many times and every time any agent's position is random, then cooperative play can be sustained as an SPNE, provided that players are sufficiently patient.

The proof is in the Appendix. What if preferences are altruistic but not impartial? With sufficient altruism, given a certain draw of time values, positive efficiency gains over FCFS are expected because some efficient swapping requests will be accepted.

Let us briefly discuss the case of incomplete information when the game is repeated finite times. With selfish preferences, it is easy to see that rejection of all requests is still optimal; therefore, sticking to their initial position is the only Nash equilibrium. Under incomplete information with non-selfish preferences, there are two sub-cases. First, if preferences are altruistic and impartial, efficient allocation is unlikely to be reached unless communication among players is allowed so that agents can inform each other about their values. When asking, agents simply reveal their values in a credible manner since incentives are perfectly aligned. Second, if preferences are altruistic but partial, almost with certainty, the efficient outcome would not be reached because, in this case, there will be conflicts among peoples' interests –and, even with communication, the full revelation of values will not occur.

2.2 Take-It-Or-Leave-It Offers (TIOLI)

In this mechanism, at any point of the trading interval before service, any player can propose to another ahead of her in the queue to swap places in exchange for money. Formally, each interaction can be characterized in tuple (i, i', j, j', q, a), where i is the index of the *requester* who is in position j and wishes to switch places with agent i', henceforth the *responder*, who is in position $j' < j$. The amount offered by the requester is denoted by $q \geq 0$, and the response, $a \in \{0, 1\}$, states whether the responder accepted or not. When the request is accepted, i (i') takes the position j' (j), and there is a transfer from i to i' equal to q. We assume that a player's utility for the whole line (trading period and service) is as described in Equation 1, adding transfers received from requests she accepted and subtracting payments she made to other players when her requests were accepted. That is:

$$u_i(j_i, \Gamma | v_i) = \mu_i - j_i v_i + \sum_{t \in \Gamma(\cdot \to i)} q_t a_t - \sum_{t \in \Gamma(i \to \cdot)} q_t a_t \tag{3}$$

where Γ represents the set of all transactions proposed in the trading period, and its restrictions $\Gamma(i \to \cdot)$ and $\Gamma(\cdot \to i)$ the sets of transactions proposed by i and to i, respectively.

Under complete information, assuming selfish and rational players, this mechanism leads to the efficient outcome where the player with the kth highest value gets the kth position in the line before service. To see why TIOLI reaches efficiency, consider first the case of a line where agents can only trade with their neighbors. Note that in any inefficient line, there must be at least one case of the following state: two players (i and i') are in neighboring positions, j and $j + 1$, respectively, where $v_i < v_{i'}$. In words, there must be two line neighbors where the person in front has a lower time value than the person behind her. Since their swap has a positive surplus and there is complete information and selfish rationality, this trade will occur. The exact price will depend on other factors impacting the bargaining process. This is sufficient to guarantee that the line will reach efficiency after a finite number of trades.

What if we allow for trades to happen with non-neighbors? The principle is similar. Let us add one element to our notation to free us from the player indexes. We will use $j_{(k)}$ to denote the current position of the player with the kth largest value, and her value will be denoted as $v_{(k)}$ —i.e., we use the order statistics of time values. In any inefficient line, there must be at least a pair of players with values $v_{(k)}$ and $v_{(k')}$ in positions $j_{(k)}$ and $j_{(k')}$ satisfying these two conditions: $j_{(k')} > j_{(k)}$ and $v_{(k')} > v_{(k)}$. Note this implies $k' < k$. However, for each order k, multiple k' orders could satisfy those conditions. The surplus of swapping the positions of these two agents can be written as:

$$w(k'|k) = \left(j_{(k')} - j_{(k)}\right)\left(v_{(k')} - v_{(k)}\right) \tag{4}$$

There are many sequences of trades in which a specific line can be dynamically sorted (different additional assumptions yield different sets Γ). It could be, for instance, that the player in position $j_{(k)}$ waits for everyone with whom she has a positive trade surplus to submit offers (generating a competition a là Bertrand) or simply accepts offers that improve her position as soon as those are received. It could be that the queue goers coordinate on competing serially for positions 1 through $n-1$ in the first round of trades, and then go back to competing for position 1 in the second round of trades, and so on.[7] It turns out it does not matter: regardless of the path, the process only ends when there are no k and k' for which $w(k'|k) > 0$, which means the line order is efficient. The exact transfers (until the efficiency order is reached) depend on the path of trades, and, therefore, it is not uniquely predicted. Interestingly, the available paths do not guarantee that the player with value $v_{(k)}$ will reach the kth position in the first round.[8] The following proposition summarizes the efficiency result.

Proposition 2. *Assuming selfish preferences and complete information, under TIOLI, the most efficient ordering of the queue can be reached in equilibrium, even if the line occurs only once.*

[7] In this case, the competition a la Bertrand, guarantees the highest payoffs for the players at the front of the line.

[8] For example, we could have the highest-value player in position 2, and the player in position 1 first accepts an offer from the player in position 3.

The proof follows the logic presented in the text above. What can we say about the size of efficiency gains and inequality? The gains from trade (relative to the initial positions) range from 0, in the case the goers are already ordered in descending value, to $\sum_{i=1}^{n}(n+1-2i)v_{(i)} > 0$ in the case that goers were ordered in ascending value initially.[9] Although this mechanism yields a reordered queue that is Pareto efficient, its effects on fairness and inequality are ambiguous. It is possible to construct examples of value distribution and initial positions that generate lower and higher inequality post-trading.[10]

What if the queue-goers only know their own value and the distribution of time value $F(v)$? Lifting the assumption of complete information naturally changes the predictions. The reordering process might be similar to the one with complete information, except the trading process of the TIOLI mechanism approximates that of simultaneous first-price private-value auctions, one for each position.

Compared to the Pareto efficient queue with complete information, the reordered queue with incomplete information is less efficient due to standard value shading during the auction. However, adding such a trading mechanism to the queue improves efficiency compared to the initial random queue since when a position is "sold", it is because the offered exceeded the value of the owner and, thus, the value of the new owner also exceeded the value of the old owner.

2.3 Social Tokens (SocTok)

This mechanism is the same as the *Swapping* mechanism, except players observe partial information regarding others' play history, namely, whether the other players were pro-social by letting others switch positions in past lines or if they mostly requested without letting others swap.

[9] This is because the total welfare of efficient sorting is $\sum_{i=1}^{n}(\mu - i \times v_{(i)})$ and the welfare of the most inefficient line is $\sum_{i=1}^{n}(\mu - (n+1-i) \times v_{(i)})$

[10] If preferences are altruistic and impartial, the multiplicity of equilibria explodes, and perfect equality can be achieved by relaxing the assumptions regarding the set of admissible transfers.

With selfish preferences, standard behavior, and under complete or incomplete information, the same results of Proposition 1 apply here. There is a family of equilibria in the finite-periods game where players do not accept any request. When some of these assumptions are lifted, the outcome could change. We postulate three possible channels.

First, it could be that this record of reputation makes indirect reciprocity possible, allowing players to be altruistic to others who, in turn, have been altruistic to a third party. Second, the information on past behavior could promote cooperation via an intrinsically worthless object (points or *social tokens*), as in Bigoni et al. (2020). Third, if part of the population is altruistic, observing behavior in past lines can serve as a signal of players' types. That is, the information provided might help players distinguish agents with altruistic preferences from selfish ones, which could sustain cooperative play within a subgroup of likely altruistic players. For these reasons, we conjecture that the SocTok mechanism should at least have the same level of efficiency as the Swapping mechanism, but it is likely to exhibit higher efficiency.

3 Experiment

We study three position-exchange rules or mechanisms, *Swapping*, *TIOLI*, and *SocTok*, each with and without unilateral communication from the requester. We deploy this experiment under a 3 $\times$ 2 full-factorial, between-group design. Each session consists of two practice rounds and 12 incentivized rounds in random matching. In each round a groups of six players endowed with a random initial position can trade positions for two minutes, before "service" and the round is over.

In the **Swapping** mechanism, at any time of the trading period, subjects can send requests to players ahead of them to switch positions, but with no monetary offer. A rational selfish subject will reject all the requests from other subjects, thus no improvement will be made to the overall efficiency. In the **TIOLI** mechanism, subjects can make a monetary offer per position along with their switch request. The requester can have only one active offer at any given time, but the responder (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, and the rest of the standing offers are automatically rejected. In the **SocTok** mechanism, no monetary transfers are allowed between the subjects. Instead, a subject earns a token if they accept someone's request and move backward, and loses a token when they move forward upon getting their request accepted.

We study each of these three decentralized mechanisms with and without unilateral communication (from the requester to responder). In half of the sessions of each mechanism, we allow participants to send a typed message along with their requests. These messages are free-form verbal communication. The responders receive and read the message together with the requests. Since the communication feature is orthogonal to the trading mechanisms, we have the six treatments referred to above: 3 mechanisms times 2 communication possibilities. That is, a treatment is defined by the pair $(mech, comm) \in \{Swap, TIOLI, SocTok\} \times \{Comm, NoComm\}$.

In the experiment, the unit of interaction is the queue. Each round corresponds to one queue, which starts with the computer grouping at random participants in groups of six. At the beginning of the round (queue), each of the six members of this group is randomly assigned a position from 1 to 6. As in the theoretical model, we do not allow for exiting or staying out of the line. In the experiment, the time values, v_i, are private, but their distribution is common knowledge.

To control for the randomness of the position assignment, we draw three random queues and rotate these realizations between rounds. To make the efficiency improvement more apparent, in all three predetermined queue starting points, most players with high service value are initially placed near the back of the line. In particular, we used the time value vectors $(2, 4, 6, 8, 10, 12)$, $(6, 2, 4, 10, 8, 12)$, and $(4, 8, 2, 6, 12, 10)$, where the first entry of the vector (read left to right) corresponds to the time value of the participant at front of the line, the second entry to the second position, and so on.[11] The arrangement of these three queues across rounds of a session makes up a "value matrix". We balance these value matrices among experimental treatments so the set of initial positions are comparable and the mechanisms have equal footing. We chose to "tilt" the initial positions towards exhibiting high inefficiency so as to generate large enough potential efficiency gains from position trading. The lines that randomly start in an efficient order, though good in real life, are less conducive to generating relevant evidence.

Once informed about their value and initial position, at any time during the round, participants 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 a subject accept a request, the two subjects (the requester and responder) switch positions, and all the other requests related to the two subjects are automatically canceled. The participants can also manually cancel their own request at any time prior to acceptance/rejection. As in the model, when person i moves forward in the line, their payoff associated with the line increases linearly (at increments of v_i). The round payoff depends on a subject's final position at the moment when service starts (the end of the round), plus/minus the subject's transfers from position exchanges when monetary transfers are allowed.

The user interface shown in Figure 1 displays from the top to the bottom: the current round information, the queue information, the actions the subjects can make, their current sent request, the requests from other subjects, their current payoff (if the current order were final), and their exchange history. For example, in this figure, we can observe that this subject is currently in round three in a TIOLI-queue with messaging possibilities. The subject is at position four and her per-position value of improving her position is 9. Though she doesn't have any active requests sent out, she has already received two requests from other subjects. From the history, we can find that the subject was originally at position

[11] The sessions were run in an incremental gain setting rather than incremental loss.

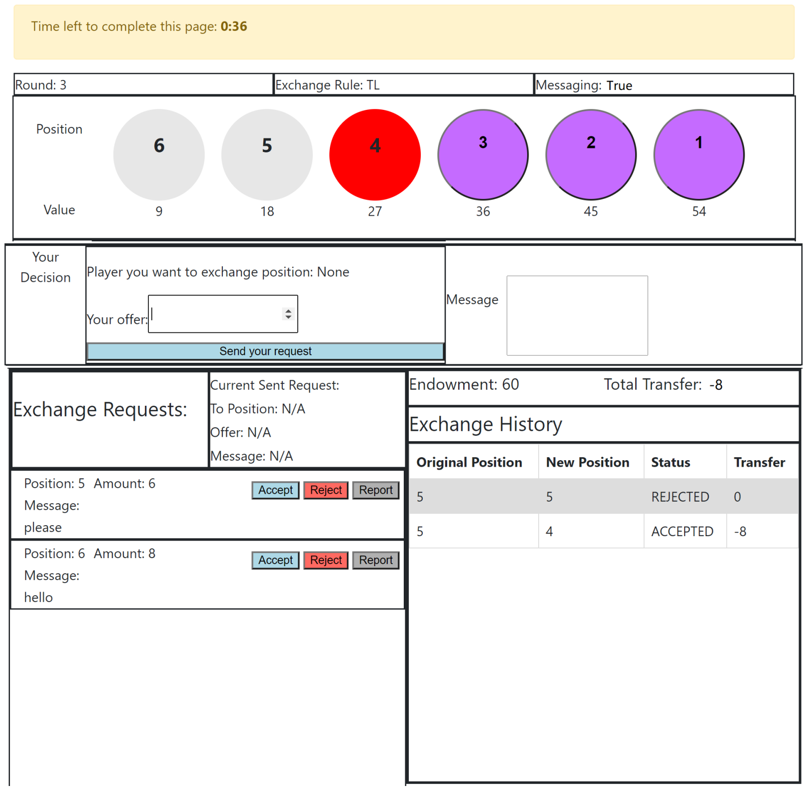


Figure 1: User interface of TIOLI treatment. *Notes*: The red dot denotes a subject's own position. Purple dots denote traders ahead of the subject in the queue. A subject's current queue messages and (round) exchange history are shown in the lower left and right sections of the screen, respectively.

5, but has traded with another subject to position 4, paying 8 experimental currency units (ECUs) in the process. To have made this move forward, she would have clicked on 4th position dot (which would have been purple), and sent a request with an offer of 8 ECUs.[12]

3.1 Procedures

We conducted 12 sessions, four per mechanism (two with communication and two without communication). In each experimental session, 12 subjects played 12 rounds each lasting two minutes (plus 2 practice rounds). At the end of each session, two random rounds were selected for payment. In each round, the subjects were divided into two 6-player groups. We applied random matching between rounds so the subjects are re-grouped between rounds. In

[12]The experiment was developed on oTree (Chen et al., 2016).

Table 1: Session, groups, and subject distribution across treatments.

Mechanism	Communication	# sessions - # groups - # subjects
Swap	Yes	2 - 48 - 24
Swap	No	2 - 48 - 24
TIOLI	Yes	2 - 48 - 24
TIOLI	No	2 - 48 - 24
SocTok	Yes	2 - 48 - 24
SocTok	No	2 - 48 - 24

total, 144 subjects interacting in 12 sessions, with 48 groups per mechanism (although only four super groups are independent). See Table 1, for the session-group-subject distribution. Subjects were recruited through the Experimental Economics Laboratory at Universidad del Pacifico on ORSEE (Greiner, 2015). We conducted the sessions online between October 2021 and March 2022. Each session lasted about an hour with an average payment of 16 PEN per subject (approximately $4 USD).

3.2 Hypotheses

Following the theoretical insights for rational selfish, we conjecture queues in the Swapping and SocTok treatments will yield no adjustments relative to the initial positions (FCFS will be respected), while there will be switching under TIOLI. This yields our baseline hypothesis:

Hypothesis 1. *(a) Under Swapping and SocTok, no request will be accepted. (b) Under TIOLI, a positive number of requests will be accepted.*

This implies the efficiency gain from the Swapping (and SocTok) mechanism should be zero relative to the initial order of the line. In contrast, the take-it-or-leave-it (TIOLI) offers can improve market efficiency as predicted by our simple model. With a complete information environment, the model predicted that the queue is fully efficient. With incomplete information (as in our experiment), we expect the queue efficiency to be between the baseline and the efficient queue. This leads to our second hypothesis.

Hypothesis 2. *(a) Under Swapping and SocTok, there is no efficiency gain relative to the initial positions (FCFS) benchmark. (b) Compared to Swapping and the initial positions, the efficiency gain is higher in the TIOLI treatment.*

However, if prosocial motivation, indirect reciprocity, or reputation are at play to some extent, the acceptance rate and the efficiency gain in the Swapping and SocTok treatments are not null.

From the model given in Section 3, in this environment, when there are no transfers, the most efficient outcome coincides with the most equitable. Conversely, when the line is the most inefficient, the line is also the most unequal. Since by design we start with highly inefficient and unequal lines, we conjecture that Swapping and SocTok treatments do not worsen inequality given that we expect little trading in those formats and if any will go in the direction of equality. However, in TIOLI we do expect trading and we expect this trading with transfers to benefit low-value players, who in our experiment tend to be in front of the line. This is because high-value players exhibit a high incentive to move ahead and at the same time they face competition in their offers to the agents at the front of the line. This implies that relatively low-value players end up with a high payment for switching; and high-value players, although able to move up, they pay a high price for doing so. In sum, inequality can be higher in TIOLI compared to the initial positions and to the efficient outcome. We summarize these conjectures in the following hypothesis.

Hypothesis 3. *Inequality of payoffs is larger under TIOLI relative to Swapping and Social Tokens.*

The communication in our experiment is *cheap talk* and is not self-committing. Some requesters could disclose their service value or a signal of being high value in the hopes of facing an altruistic or efficiency-oriented peer and being able to move forward. However, deception is also possible. Being cheap talk, we hypothesize that allowing messages from the requester to the responder will not improve queue efficiency.

Hypothesis 4. *Unilateral communication from requesters to responders does not change behavior, nor improve queue efficiency.*

4 Results

We first offer a descriptive overview of the experimental data, then we describe and present our main results emerging from standard regression analysis, and finally offer a brief discussion on the content of free-form communication.

4.1 Descriptive Statistics

In Table 2 we present the main descriptive statistics of our experimental results. As mentioned before, we have 96 queues per mechanism (48 with communication and 48 without). The unit of interaction in our experiment is the request. We observe that while Swapping exhibits almost 14 requests in each line, TIOLI and Social Tokens exhibit fewer requests per line (11.6 and 11.4, respectively). It is interesting to note that, as a reference, if every person would request everyone ahead of them once (as predicted in the model section above), the total number of requests would be 15.

The main difference we find is the acceptance rate. While Swapping exhibits an acceptance rate of only 9.4%, TIOLI has one of 21%, and Social Token has an acceptance of 11.4%. Similarly, when communication is allowed, the number of messages requesters send varies widely across mechanisms. In Swapping, on average there are 7.9 messages per line, in TIOLI, only 2 messages and on Token there are 5.6 messages. That is, the message rate (# message/#requests) is much lower in TIOLI compared to the other mechanisms: Swapping 68%, TIOLI 12%, and Token 60%. In TIOLI, the average monetary offer is 6.7 ECUs per position but the average accepted offer is 9.1 ECUs. There is not much difference with or without communication.

Table 2: Descriptive Statistics

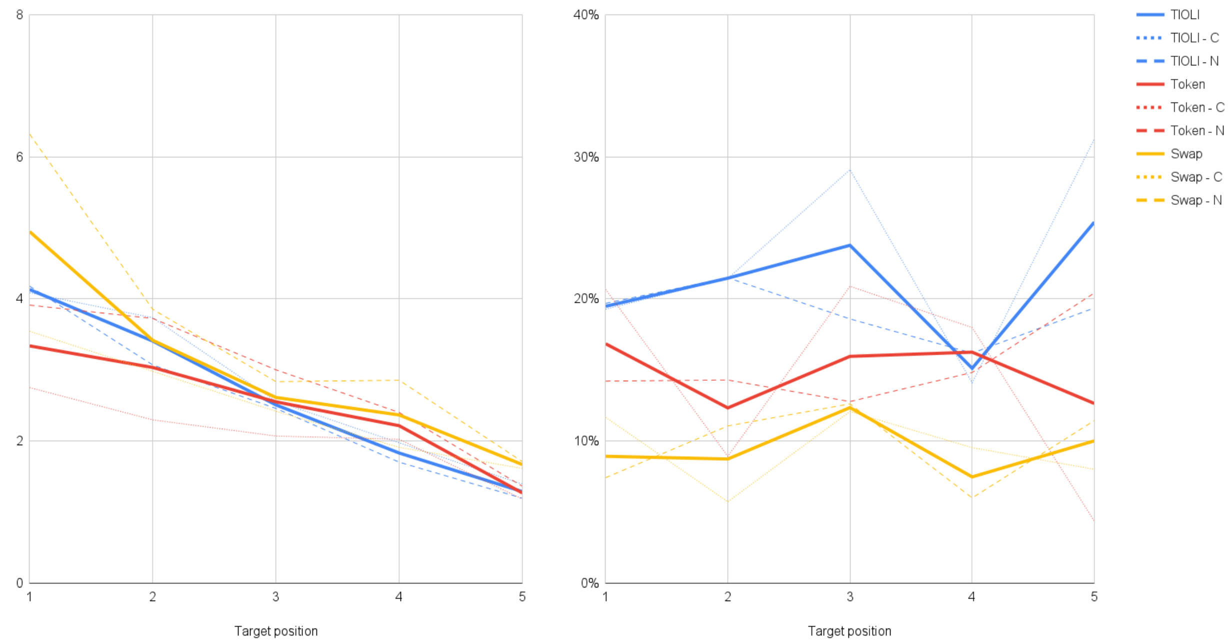
	Swapping			TIOLI			SocTok		
	Total	Comm	NoCom	Total	Comm	NoCom	Total	Comm	NoCom
Num. of queues	96	48	48	96	48	48	96	48	48
Num. of switch requests	13.9	11.5	16.3	11.6	12.0	11.1	11.4	9.3	13.5
Acceptance rate (%)	9.4	9.8	9.1	21	22	20	15.3	15.9	14.9
Average offer				6.7	6.4	7.1			
Average accepted offer				9.1	8.9	9.3			
Num. of messages		7.9			2.0			5.6	
Efficiency gain (%)	6.4	7.6	5.1	34.4	36.7	32.1	7.6	5.2	10.1
Average payoff per round	80.5	80.6	80.3	83.5	83.7	83.2	80.6	80.3	80.8
Inequality index	1.24	1.28	1.21	1.36	1.35	1.37	1.28	1.23	1.33

Notes: Except for the number of queues, all variables are calculated as averages per queue (session-group-round). The number of switch requests is simply how many times in the round any participant requested a position switch to another participant. The acceptance rate is the percentage of times that request was accepted. The average offer only applies to TIOLI and it is the average offer in ECUs. The average accepted offer is the same as the previous statistic but is now conditional on the request being accepted. The number of messages refers to the number of typed messages sent from requesters to responders. The efficiency gain is defined in the text and mainly refers to the percentage of the maximum welfare increase that is realized in the position-trading process. The inequality index is the ratio of the standard deviation of the final payoff over the standard deviation of the initial payoffs.

Next we analyze the efficiency gains, defined as in Equation 5, where W_{initial}, W_{final}, and $W_{\text{efficient}}$ denote, respectively, the sum of payoffs with the initial positions (as if the first-come-first-serve principle would have been followed), with the observed final positions and with the efficient allocation of positions.

$$\text{Efficiency Gain} = \frac{W_{\text{final}} - W_{\text{initial}}}{W_{\text{efficient}} - W_{\text{initial}}} \tag{5}$$

That is, the definition of efficiency gain given in Equation 5 takes a value of 0 if the final welfare is the same as the initial, and it takes a value of 1 if the line reaches the maximum possible welfare. We observe significant increases in the efficiency of the line (+34.4%) in TIOLI, and, to a much lesser degree, in Swapping (6.4% gain) and Social Token (7.6%).



(a) Number of requests per target position (b) Acceptance rate

Figure 2: The frequency of requests by responders' position and acceptance rate.

Communication seems to slightly improve efficiency in Swap and TIOLI but the opposite happens in Social Token. We will get back to this in the regression analysis. We also analyze the payoffs per round and found that Swap and Token exhibit similar final payoffs while TIOLI presents a higher payoff, due to its improvement in efficiency, as expected. Communication availability makes not much difference in this dimension. Lastly, in the same table, we report a measure of inequality: the ratio of the standard deviation of round payoffs over the standard deviation of the initial payoffs. That is, how much more inequality there is relative to the FCFS benchmark? We find that as conjectured, TIOLI's increase in efficiency comes with an increase in inequality.

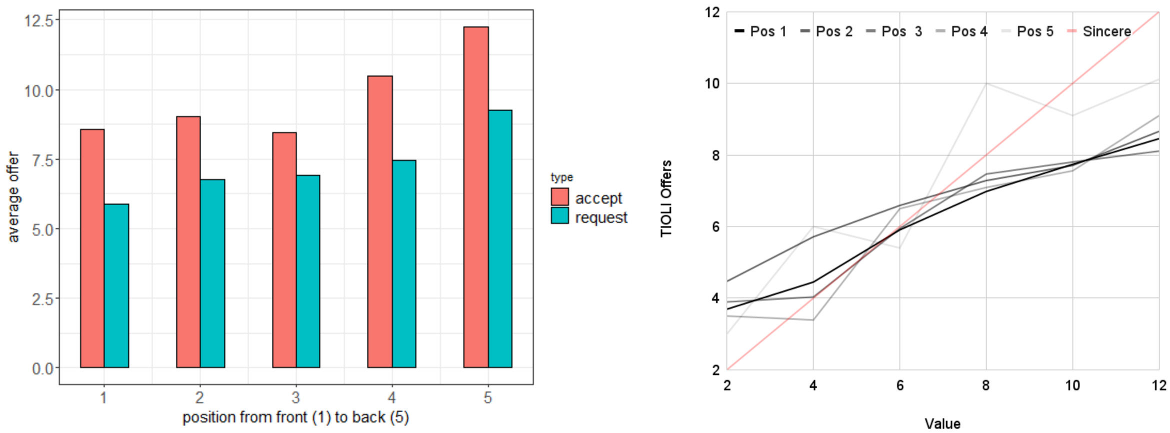
We then inspect the frequency of requests by the target (responder's) position. If everyone were to request once for each position ahead of them, we would have five requests for the first position, four for the second, and so on, until we have one request for position five and zero for position six. This would total 15 requests. As shown in Figure 2 Panel (a), the frequency of requests is not that different from this theoretical prediction that assumes queue goers would settle positions competeting for all spot. We find that for all target positions Swap exhibits slightly more requests, then SocTok, then TIOLI; but overall not much difference beyond target position 1 appears. We do find that the main difference (as reported above) lies in the acceptance rate. For nearly all positions, TIOLI has a higher acceptance rate, then Social Token and last Swap–which exhibits an acceptance rate half of TIOLI's. Interestingly, this order holds approximately for all target positions in Figure 2 Panel (b).

Offer behavior in TIOLI

Let us now discuss offer behavior in TIOLI. We need this separate inspection since this is the only treatment with monetary transfers.

As we can see in panel (a) of Figure 3, the average offers by target position go from approximately 6 ECUs for position 1 to 9 ECUs for position 5. This increasing pattern is due to the fact that the lines we study have higher-value people at the back of the line and consequently the offers for the last positions are made on average by higher-value participants. Naturally, there is a similar pattern that occurs in the accepted offers that go from 8.7 ECUs for position 1 to 12.3 ECUs for position 5.

In every request under TIOLI, if the requester has a higher time value than the responder (positive surplus), the overall efficiency increases when the request is accepted, and both players earn money if the offer price is between their time values. However, not all encounters are efficient nor do all transactions make money for both players. In 67.8% of the exchanges both players earn some profit, only the requester loses money in 27.0% of the exchanges, and only the responder loses money in 3.4% of them. Finally, in 1.7% of the exchanges, both players lose. Here, we concentrate on the first two cases – since they amount to 95% of all encounters of this kind. First, we find evidence that the responders tend to have a higher share of surplus in the TIOLI treatment. This is predicted by the theoretical insights given the structure of the game: simultaneous first-price auctions where competition among bidders leaves the seller with most of the surplus. In Appendix C, we document that the



(a) All requests

(b) Accepted requests

Figure 3: The frequency of requests by responders' position.

responders often obtain 90% share of the normalized surplus or more, and the requester tends to earn little.

Second, we look into cases where the requester submits an offer that could make them lose money. The user interface displays a pop-up with a warning message stating they could decrease their material payoff; as such, these losses are not resulting from miscalculation. From panel (b) of Figure 3 we can see that, on average, it is low-value participants that tend to send unprofitable offers (offers that are higher than the 45-degree line). Also, from Appendix C, in Figures 4 and 5, we see that these types of requests are sent by a few subjects (7 subjects contribute 62% of the total such exchanges, 39 out of 63 requests). The majority of these subjects do not accept any offer after the exchange; that is, they rarely earn the money back later and tend to lose money in the round.[13]

There are two main potential explanations: players are trying to speculate or they attach intrinsic value to the action of moving ahead or being at the front of the line (similar to what the literature has referred to as experiencing "the joy of winning").[14] The evidence is not

[13]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.

[14]Learning cannot explain the irrationality here, since a significantly number of irrational trades happened in later rounds.

conclusive: since lower-value players are on average overbidding, we conjecture that the intention to speculate was in place. However, the fact that the players were not exchanging nor making a profit later could be because they had the incorrect belief they could resell those positions (and they could not) or because they indeed put an intrinsic value on winning positions.

4.2 The effects of different position-switching mechanisms

We conduct a regression analysis to provide statistical inference on the insights provided in the descriptive analysis above. Since we find no clear indication that unilateral communication had any impact, we first focus on analyzing the trading mechanisms alone. In the next subsection, we incorporate communication into the analysis.

Table 3 reports treatment effects on the following outcomes: (1) the number of position-switching requests; (2) the acceptance rate of those requests; (3) the average round payoff; (4) efficiency gain relative to the initial distributions of positions; (5) payoffs inequality; and (6) the number messages (when allowed). In the regressions, the outcomes are defined at the queue level (session-group-period) and we used as baseline (comparison group) the initial state of the line, which is what would happen if the first-come-first-serve (FCFS) principle was respected. For this reason, we can directly interpret the coefficients as deviations with respect to FCFS. At the bottom of the table, we present the t-tests comparing each mechanism against the other.

We find that in all mechanisms, the number of requests is clearly above zero (14 for Swap and 11.5 for TIOLI and SocTok). Also, Swapping exhibits a statistically higher number of requests than the other two mechanisms, which in turn are not different from each other. This is consistent with the fact that in Swapping, the cost of requesting and getting accepted is zero in monetary and non-monetary resources. Regarding the acceptance rate, we again find that the rate in all mechanisms are above zero (10% for Swap, 16.6% for Social Token, and 21 % for TIOLI). This increasing ordering is statistically significant

$0 <$ Swap $<$ Social Token $<$ TIOLI. This is consistent with the fact that in Swapping, there is no material nor reputational benefit to accepting a request if the agent is selfish, while, between Token and TIOLI, the latter does have a material benefit. This generates our first finding.

Result 1. *The acceptance rates in all of the three mechanisms are strictly positive. The acceptance rates of the three mechanisms are all statistically different, with TIOLI exhibiting the largest rate, followed by SocTok, which, in turn, has a larger rate than Swapping.*

Interestingly, as conjecture for cases that depart from the standard model, in the Social Token mechanism, whether a player accepts a request seems to depend positively on the requester's reputation. Table 2 in the Appendix shows the logit regression that models the acceptance decision as dependent on the requester's number of tokens, using the data solely from the Social Token mechanism. The associated coefficient is positive and significant –i.e., players are likelier to accept requests from people with more tokens. This suggests that people respond to reputation or are motivated by indirect altruism.

For efficiency gains, all mechanisms exhibit a positive gain relative to FCFS: 0.06 for Swap, 0.8 for Social Token, and 0.34 for TIOLI. The corresponding t-tests show that only TIOLI exhibits a higher average payoff relative to the other two mechanisms ($0 <$ Swap $\sim$ Social Token $<$ TIOLI). All the mechanisms improve efficiency relative to the original order of the line, but only TIOLI exhibits a substantial improvement. The other two mechanisms show a modest efficiency gain. This pattern also applies to round payoffs, as seen in the column before efficiency in Table 3. This is summarized in our next result.

Result 2. *The efficiency gains of the three mechanisms are strictly positive. TIOLI presents the largest gain, followed by SocTok, and finally Swapping.*

Finally, we analyze the impact of the different mechanisms on inequality. We find that all mechanisms increase inequality relative to the initial positions but not to the same degree. TIOLI, as hypothesized, increased inequality more than the other mechanisms. However,

Table 3: Regression Analysis - Only Switching Mechanisms

	Numb. of requests	Acceptance rate	Average Payoff	Efficiency	Inequality
Swap	13.94***	0.0998***	0.674***	0.0636***	0.245***
	(0.996)	(0.00954)	(0.163)	(0.0112)	(0.0439)
Token	11.44***	0.166***	0.809***	0.0763***	0.279***
	(0.467)	(0.0134)	(0.183)	(0.0138)	(0.0524)
TIOLI	11.55***	0.217***	3.674***	0.344***	0.358***
	(0.394)	(0.0133)	(0.290)	(0.0267)	(0.0636)
Roundnumber	0.262***	0.000788	0.0472*	0.00453*	-0.000850
	(0.0879)	(0.00145)	(0.0238)	(0.00212)	(0.00651)
Intercept	-2.224**	-0.00669	79.38***	-0.0385	1.007***
	(0.753)	(0.0123)	(0.220)	(0.0181)	(0.0554)
R-squared	0.486	0.383	0.397	0.422	0.080
Observations	384	384	384	384	384
Token-swap t-test	-2.5***	0.066***	0.135	0.013	0.035
	(1.092)	(0.016)	(0.212)	(0.018)	(0.068)
TL- swap t-test	-2.385***	0.117***	3***	0.28***	0.113*
	(1.063)	(0.016)	(0.309)	(0.029)	(0.077)
TL-Token t-test	0.115	0.051***	2.865***	0.268***	0.078
	(0.597)	(0.019)	(0.32)	(0.03)	(0.082)

Note: * $p < .1$, ** $p < .05$, *** $p < .01$ (standard errors in parentheses).

these differences are only mildly significant and only for TIOLI's inequality relative to that of Swapping.

Result 3. *Inequality increases relative to the initial positions are strictly positive under all mechanisms. However, the inequality increases under Swapping and Social Tokens are smaller than under TIOLI.*

4.3 The effects of communication

We now analyze the impact of communication on the same outcomes and within the studied mechanisms. As explained earlier, communication in our experiment occurs only from a requester to a responder in a better position in the line. In Table 4, we report the estimates from a similar specification as seen in Table 3, except we now add the communication interactions. The main finding is that communication has little to no impact on the key outcomes. In the intermediate outcomes, communication reduces the number of requests in Swap and SocTok. We argue this is because people feel they need to send some message along with the request, which might be perceived as an effort cost.

Interestingly, we also confirm that participants send more messages in Swap, followed by Token, and, last, by TIOLI. In the main outcomes, only Social Token exhibits slightly less efficiency when communication is allowed than in the environment without communication. This effect, however, is significant only at the 10% level. Besides the differential effect on the number of requests, being in the environment with or without communication does not alter qualitatively the contrast among mechanisms we documented in the previous subsection. This is summarized in our next result.

Result 4. *Unilateral communication from a requester to a responder ahead in the line reduces the frequency of requests in Swap and Token but does not impact queue efficiency or inequality.*

Table 4: Regression Analysis - Incorporating Communication

	N of requests	Acceptance rate	N of messages	Average Payoff	Efficiency	Inequality
Swap	16.33***	0.102***	7.854***	0.549*	0.0513***	0.209***
	(1.865)	(0.0145)	(0.471)***	(0.201)	(0.0146)	(0.0631)
Token	13.54***	0.169***	5.562***	1.063***	0.101***	0.330***
	(0.736)	(0.0200)	(0.407)	(0.268)	(0.0220)	(0.0760)
TIOL	11.15***	0.205***	1.979***	3.451***	0.321***	0.368***
	(0.407)	(0.0175)	(0.371)	(0.400)	(0.0382)	(0.0911)
$Swap \times Comm$	-4.792**	-0.00399		0.250	0.0246	0.0713
	(1.929)	(0.0191)		(0.276)	(0.0222)	(0.0878)
$Token \times Comm$	-4.208***	-0.00621		-0.507	-0.0494*	-0.102
	(0.810)	(0.0269)		(0.319)	(0.0271)	(0.105)
$TIOL \times Comm$	0.813	0.0249		0.444	0.0459	-0.0211
	(0.764)	(0.0266)		(0.554)	(0.0532)	(0.128)
RoundNumber	0.262***	0.000788	0.128**	0.0472**	0.00453**	-0.000850
	(0.0832)	(0.00145)	(0.0513)	(0.0239)	(0.00212)	(0.00653)
Intercept	-2.224***	-0.00669	-1.090*	79.38***	-0.0385**	1.007***
	(0.713)	(0.0123)	(0.441)	(0.221)	(0.0181)	(0.0555)
R-squared	0.528	0.385	0.611	0.404	0.429	0.084
Observations	384	384	384	384	384	384
T-tests without communication						
Token-swap	-2.792	0.067***		0.514	0.05*	0.121
	(2)	(0.025)		(0.312)	(0.026)	(0.099)
TL- swap	-5.188***	0.103***		2.903***	0.27***	0.159
	(1.904)	(0.023)		(0.43)	(0.041)	(0.111)
TL-Token	-2.396***	0.036		2.389***	0.22***	0.038
	(0.831)	(0.027)		(0.466)	(0.044)	(0.119)
T-tests with communication						
Token-swap	-2.208***	0.065***	-2.292***	-0.243	-0.024	-0.052
	(0.614)	(0.022)	(0.616)	(0.283)	(0.023)	(0.094)
TL- swap	0.417	0.132***	-5.875***	3.097***	0.291***	0.067
	(0.824)	(0.024)	(0.593)	(0.445)	(0.041)	(0.108)
TL-Token	2.625***	0.067**	-3.583***	3.34***	0.315***	0.119
	(0.741)	(0.027)	(0.543)	(0.438)	(0.04)	(0.115)

Note: * $p < .1$, ** $p < .05$, *** $p < .01$ (standard errors in parentheses).

Table 5: Message Features.

	Swapping	Social Token	TIOLI
Total number of messages	377	267	95
(1) With offer	0.65	0.65	0.87
(2) Explicit request	0.42	0.57	0.09
(5) Persuasive / emotion	0.38	0.22	0.23
(3) Mentions surplus	0.21	0.03	0.18
(4) Mentions own value	0.16	0.06	0.00
(6) Deception attempt	0.06	0.30	0.01

Finally, we analyze the full set of in-queue typed messages. We classify messages according to six non-mutually-exclusive features: (1) "With offer": when the message includes an explicit mention of the offer in numeric/monetary terms; (2) "Explicit request": when the message explicitly states a request; (3) "Mention surplus": when the message mentions a gain for the message recipient associated with accepting the offer; (4) "Mention own value": when the message mentions the own potential gain of the transaction; (5) "Persuasive": when the message appears to be persuasive in a salesperson or effective/emotional sense; and (6) "Deception": when the message is considered to attempt deception or to reflect a misunderstanding;

In Table 5, each column refers to a specific category that the messages contain. Note that a message can belong to multiple categories or none. The numbers show the fraction of messages that exhibit the feature. Most messages contain an explicit request or an offer, clearly showing the requesters' intentions. However, the requester often fails to send informative and convincing requests. In the Swapping and SocTok treatments, the responder's acceptance implies a material loss. We found that the requester in these institutions tends to be more persuasive and deceptive. On the contrary, in the TIOLI treatment, the requester tends to either say nothing (fewer messages) or discuss the payment or the surplus gains.

5 Conclusions

The first-come-first-serve principle is accepted as being fair even though inefficient. Centralized mechanism alternatives have been presented to improve efficiency yet face challenges from complexity or for being less practical. We ask the question of what simple decentralized mechanisms could improve queue efficiency. These mechanisms have been under-researched to the best of our knowledge.

We test three simple decentralized mechanisms (swapping, Take-it-or-Leave-it, and Social Token) in settings with and without the ability to communicate. We find that people make more requests in Swapping than in SocTok and more in SocTok than in TIOLI. The acceptance rates of all three mechanisms are strictly positive. Yet, TIOLI exhibits the largest rate, followed by SocTok and Swapping, in that order. In terms of efficiency, all treatments exhibit improvements over the initial efficiency of the line. However, Swap and SocTok do so to a significantly smaller degree than TIOLI (6-7 percentage points of the total efficiency possibilities). The TIOLI mechanism increases efficiency by 34 percentage points of the total possible gains. Although the lines are more efficient in TIOLI, most of the "gains from trade" go to the players who own the front positions, which hurts equity.

We also find that whether unilateral communication is allowed or not, has little to no effect on the primary outcomes we study: efficiency and inequality. We find that, in a few exchanges, some requesters propose unprofitable offers. We argue there is mixed evidence of speculation intentions or intrinsic value for (getting) the front positions (joy-of-winning).

References

Afeche, P. and Mendelson, H. (2004). Pricing and priority auctions in queueing systems with a generalized delay cost structure. *Management science*, 50(7):869–882.

Afeche, P. and Pavlin, J. M. (2016). Optimal price/lead-time menus for queues with customer choice: Segmentation, pooling, and strategic delay. *Management Science*, 62(8):2412–2436.

Allon, G. and Kremer, M. (2018). Behavioral foundations of queueing systems. *The handbook of behavioral operations*, 9325:325–366.

Andreoni, J. and Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of public economics*, 95(7-8):513–520.

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.

Chatterjee, K. and Samuelson, W. (1983). Bargaining under incomplete information. *Operations research*, 31(5):835–851.

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.

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.

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

Katok, E., Leider, S., and Donohue, K. (2018). The handbook of behavioral operations.

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

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.

Oberholzer-Gee, F. (2006). A market for time fairness and efficiency in waiting lines. *Kyklos*, 59(3):427–440.

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.

Rosenblum, D. M. (1992). Allocation of waiting time by trading in position on a g/m/s queue. *Operations Research*, 40(3-supplement-2):S338–S342.

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.

Shunko, M., Niederhoff, J., and Rosokha, Y. (2018). Humans are not machines: The behavioral impact of queueing design on service time. *Management Science*, 64(1):453–473.

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.

Appendix

A Proposition 1 Proof

Proposition 1: *In the environment described above, under complete information:*

(a) If preferences are selfish and the queue is repeated for a finite number of periods, the only Nash equilibrium is not to accept any swapping request.

(b) If preferences are altruistic and impartial, cooperative play is a Nash equilibrium regardless of the number of periods the queue is played.

(c) If the queue is repeated infinitely many times and every time any agent's position is random, then cooperative play can be sustained as an SPNE, provided that players are sufficiently patient.

Proof. (a) If the game is played once, there is no incentive for a selfish individual to accept a swap (it only hurts her utility). This item of the proposition follows from the standard results on finite-horizon games with unique Nash in the stage game.

(b) Everyone has the same preferences. Therefore, it suffices to show that swapping the positions of any pair of players, i and i', where i has a higher time value but is somewhere behind i', increases the utility. Because $f(.)$ is strictly increasing, moving player i' backward in the line negatively impacts the utility, and the impact of moving player i forward in the line on the utility is positive. Furthermore, by the symmetry imposed by the impartiality assumption, the latter is a stronger impact than the former. In terms of behavior, this implies that any request to swap from a higher-value person will be accepted. Therefore, any person who sees a lower-value individual ahead in the line is incentivized to ask for a swap. These requests stop when the line is sorted efficiently.

(c) This item follows the standard folk theorem. Since the cooperative outcome gives, in expectation, the highest payoff for any player compared to any non-cooperative play, it can be sustained if players are sufficiently patient. For example, everyone sorts themselves into the efficient line, and if anyone refuses, then no one else requests any swaps nor accepts them anymore, ever.

□

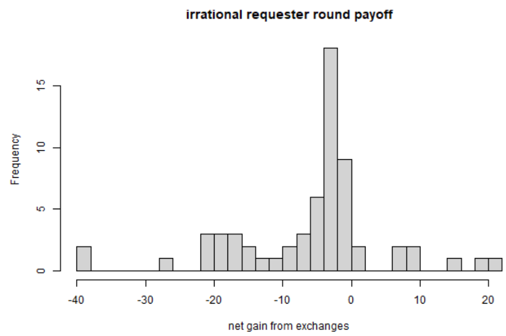
B Auxiliar Statistics (Rounds 3-12)

Table 1: Descriptive Statistics (Rounds 5-14)

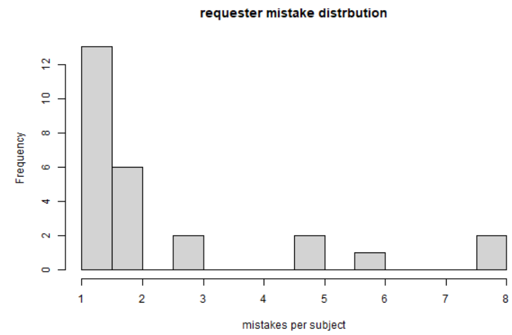
	Swapping			TIOLI			Social Token		
	Total	Comm	NoCom	Total	Comm	NoCom	Total	Comm	NoCom
Num. of queues	72	36	36	72	36	36	72	36	36
Num. of switch requests	14.4	11.8	17.1	12.0	12.9	11.1	11.9	9.8	14.1
Acceptance rate (%)	9.7	9.2	10.1	21.7	23.1	20.2	15.2	17.1	13.8
Average offer				6.7	6.3	7.2			
Average accepted offer				9.2	9.0	9.4			
Num. of messages	8.5	8.5	0.0	2.0	2.0	0.0	6.1	6.1	0.0
Efficiency gain (%)	7.3	8.3	6.2	38.6	43.4	33.8	8.2	5.0	11.4
Average payoff per round	80.5	80.6	80.4	83.9	84.4	83.4	80.6	80.3	81.0
Inequality index	1.26	1.29	1.24	1.38	1.40	1.36	1.27	1.21	1.34

Notes: We only consider rounds 3 to 12. Except the number of queues, all variables are calculated as averages per queue (session-group-round). Number of switch requests is simply how many times in the round any participant requested a position switch to another participant. The acceptance rate is the percentage of times that request was accepted. The average offer only applies to TIOLI and it is the average offer in ECUs. The average accepted offer is the same as the previous statistic but now conditional on the request being accepted. The number of messages refers to the number of typed messages sent from requesters to responders. Efficiency gain is defined in the text and mainly refers to the percentage of the maximum welfare increase that is realized in the position-trading process. The inequality index is the ratio of the standard deviation of the final payoff over the standard deviation of the initial payoffs.

C Offer Behavior in TIOLI



(a) Round payoff for the irrational requester.



(b) The distribution of requesters making mistakes.

Figure 4: Irrational requesters

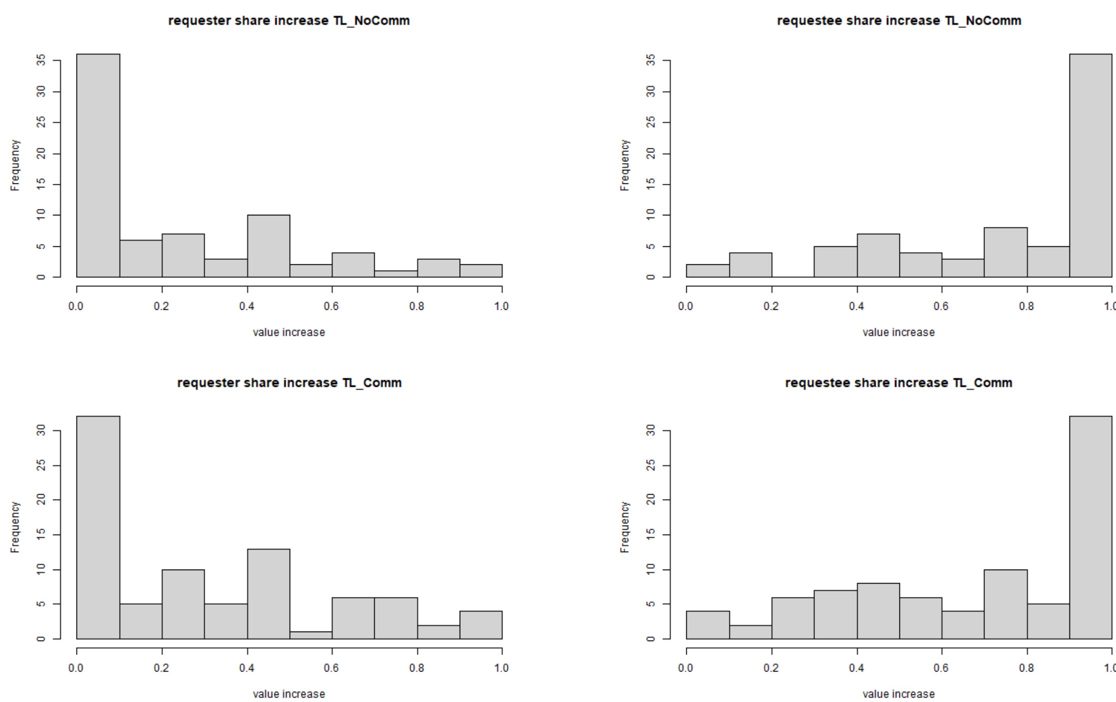


Figure 5: value increase per position from the accepted exchange.

D Additional Results in SocTok

	(1)	(2)
(Intercept)	-1.4728***	-0.8577***
	(0.0000)	(0.0018)
requester token	0.2014***	0.1924***
	(0.0000)	(0.0000)
requester value		-0.0890***
		(0.0029)
responder value		-0.0107
		(0.6867)

Table 2: Logit regressions at the request level. The dependent variable is a dummy variable that equals to 1 when the request is accepted.