Regulation of Organ Transplantation and Procurement: A Market Design Lab Experiment

Alex Chan[1] and Alvin E. Roth[2]

[1]Harvard University

[2]Stanford University

July 17, 2023*

Abstract

We conduct a lab experiment that shows current rules regulating transplant centers (TCs) and organ procurement organizations (OPOs) create perverse incentives that inefficiently reduce both organ recovery and beneficial transplantations. We model the decision environment with a 2-player multi-round game between an OPO and a TC. In the condition that simulates current rules, OPOs recover only highest-quality kidneys and forgo valuable recovery opportunities, and TCs decline some beneficial transplants and perform some unnecessary transplants. Alternative regulations that reward TCs and OPOs together for health outcomes in their entire patient pool lead to behaviors that increase organ recovery and appropriate transplants. (*JEL* Codes: C92, D47, I18)

Keywords: Organ Transplantation, Lab Experiment, Market Design, Regulation, Healthcare

*Email addresses: achan@hbs.edu (A.C.); alroth@stanford.edu (A.E.R.). This study was approved by the Institutional Review Board at Stanford University (IRB-44868). This paper also benefited from comments from participants in a number of conferences and seminars (NBER Market Design Working Group, Economic Science Association North America Annual Meeting, Stanford University), and from referee and editorial suggestions.

1 Introduction

Deceased donor organ transplantation is in turmoil. Over 110,000 Americans are waiting for an organ transplant and over 5,770 died waiting for an organ in 2019. In the same year, 5,957 recovered organs were discarded. There are multiple Congressional investigations, and active proposals to improve regulations of deceased organ procurement and transplantation, put forth by non-governmental agencies like the Federation of American Scientists and the National Academies of Sciences, Engineering, and Medicine (US Congress (2021), Federation of American Scientists (2021), National Academies of Sciences, Engineering, and Medicine (2018), National Academies of Sciences, Engineering, and Medicine (2021)). These investigations and proposals look for ways to increase organ transplantation, by increasing recoveries and reducing discard of already recovered organs. The bulk of this discussion had been focused on the regulation and performance measurement of organ procurement organizations (OPOs). Separate discussions have occurred regarding performance monitoring of transplant centers (TCs). Among steps taken to date are a major policy change to remove 1-year graft and patient survival standards as a TC re-certification requirement by the Centers for Medicare and Medicaid Services (CMS) (Centers for Medicare and Medicaid Services (2019)). Although OPOs and TCs interact with each other, these regulatory discussions largely considered them separately.[1] Since the incentives and opportunities facing OPOs and TCs are intertwined, such fragmented approaches might be inefficient.

A deceased organ transplantation starts with a death. A hospital must contact its local Organ Procurement Organization (OPO) about every patient's death. This has been required by law since 1998 (Health Care Financing Administration (1998), Siminoff et al. (2001)). The OPO will then obtain information to assess whether the deceased is eligible for organ donation. Given limited resources, OPOs have to prioritize which cases to pursue and which ones to pursue first. There are 57 OPOs in the US, each exclusively responsible for recovering organs in their designated donor service area (Mone (2002)).[2] These OPOs are not-for-profit organizations.[3]

[1]The American Society of Transplantation is working with the American Society of Transplant Surgeons to review performance metrics but only focused on TCs (Phend (2020)), while groups like Federation of American Scientists and ORGANIZE are focused on OPOs (Rosenberg et al. (2020)).

[2]While Mone (2002) lists 58 OPOs, two OPOs in New England merged since.

[3]Held et al. (2021) found that OPOs generate millions of dollars in "profits" (or revenue in excess of costs, sometimes referred to as net assets for not-for-profit organizations) annually and hold tens

Should an organ be recovered by an OPO, deceased organ allocation would happen through the Organ Procurement and Transplantation Network (OPTN) platform, and offers of the recovered organ will be made to patients on waitlists maintained by TCs based on pre-determined priority rules. TCs do not control the prioritization of patients on their waitlists. What TCs have the freedom to do is to determine whether they want to reject an organ offered for a particular patient on their waitlist, or advise that patient to accept the organ.[4] In practice, the decision whether or not to accept a kidney when it becomes available to a patient is made predominantly by their transplant surgeon (Solomon et al. (2011)).[5] TCs are typically a part of a larger for-profit or not-for-profit health system or hospital that relies on reimbursements (e.g., from performing transplants) as a source of income.

There are two prominent areas where mis-incentives might drive undesirable OPO and TC behavior. First, TCs are penalized (by the OPTN Membership and Professional Standards Committee [MPSC]) for adverse health outcomes among patients who the TCs transplant, but not for those who remain untransplanted on the waitlist. Specifically, TCs that fail to meet the minimum standards for one-year post-transplant patient and graft survival can face severe penalties. So, TCs might be incentivized to restrict transplants to healthier patients using higher quality kidneys even when some sicker patients could have benefited from a lower quality kidney, that is instead discarded. Second, while the TCs are incentivized to be selective with kidneys offered to them by OPOs, the OPOs are incentivized to avoid recovering kidneys that will be declined by TCs. That could induce OPOs to forgo opportunities to recover organs from donors who are unlikely to yield readily accepted organs, especially deceased donors who fall along the margins of what qualifies the deceased as eligible donors.

We use a lab experiment to investigate how a holistic approach that focuses on align-

of millions of assets. While not-for-profit in nature, executive pay at OPOs is often comparable with healthcare sector executive pay and management is often incentivized to increase OPO profits for both the solid organ (kidneys, livers, hearts, lungs, etc.) recovery operations and for the more financially lucrative tissue recovery business.

[4]TCs can manipulate who is on their waitlist but for simplicity we speak of waitlist and the transplant-eligible patient pool interchangeably here. However, we recognize that a patient pool measure that is not easily manipulable by TCs will be necessary for actual policy design: see the concluding discussion. Also, for convenience, we speak of the waitlist as if it represents the whole pool of transplant-eligible patients. However, we will note in the conclusion that a non-manipulable measure of transplant-eligible patients remains to be constructed.

[5]Often patients are not told when an organ is declined on their behalf, even after the fact Husain et al. (2019)). This is also the reason why previous efforts in the economics literature to model the kidney accept/reject decision have modeled it as a surgeon decision alone (see for example, Howard (2002)).

ing OPO's and TC's interests to promote population health might offer a better design for organ transplant regulation. We focus on procurement and transplantation of kidneys. We show that moving from the current fragmented regulation to holistic regulation where TCs and OPOs are jointly rewarded for health improvements for all patients (transplanted or waitlisted but not transplanted) can shift the organ recovery and transplantation rates. We also look at the nature of these shifts to see whether they benefit patients on average and whether the forgone recovery or transplantation opportunities could have benefited patients. While no actual transplantation takes place in our experiment, the decision architecture approximates the environment that OPOs and TCs face. Besides providing clean identification and counterfactuals, our experiment also lets us observe outcomes analogous to important outcomes that are typically unobservable in the wild. For instance, we are generally unable to observe unrecovered kidneys, or whether discarded organs could have benefited some transplant candidates, but we can observe the analogous outcomes in our experimental setting.

We adopt a simplified representation of the OPO-TC dynamic in a lab setting to represent the interactions between one OPO and one TC where the former decides whether to recover two kidneys from a deceased donor, and the latter (if the kidneys are recovered) gets the option to perform kidney transplants with two kidney transplant candidates (one sicker and riskier, and one healthier and safer).

To simulate the effects of different incentive systems, we compare player behavior across *status quo* and *holistic* regulation conditions. The conditions differ in how payoffs are determined for the players. The instructions to subjects are stated in abstract terms (involving red and blue balls in urns or jars), not in terms of patients and organs. The risks associated with transplanting a particular kidney into a particular patient are based on the risk profiles of both patient and kidney.

Comparing the average behavior of the players across the conditions, we can assess whether there is evidence for shifts in the behaviors representing organ recovery and discards under alternative incentive schemes. Second, we can review the true qualities of the "kidneys" and "patients" to assess whether some forgone recoveries and/or discards could have improved the underlying probabilities of patients getting "good health outcomes." Our experiment also allows us to evaluate the policy impact on the health outcomes in our experiment. We deliberately use the term "good health outcome" to broadly represent desired health outcomes including not just one-year graft survival but outcomes like more patient quality-adjusted life years (QALYs) or a broader set of outcomes that include improved QALYs across multiple stages of the patients' life-cycle of care.

We have several main findings. First, more kidneys are recovered and offered but discards do not rise under the holistic regulation condition, so more kidneys are transplanted. This suggests that the incentives under the current regulatory regime might be inducing OPOs to under-recover kidneys and TCs to over-discard recovered kidneys. Second,between conditions there are significant differences in recovery and discards that are "missed opportunities" that could have helped some patients. Third, there are gains in good health outcomes under the holistic regulation relative to the status quo. Health benefits are particularly pronounced for the sickest and healthiest transplant candidates - there are more helpful transplants for sick patients and fewer inappropriate transplants for healthier patients.

The problems and limitations of regulatory enforcement of performance standards in organ transplantation using simple metrics have been widely debated in the clinical community.[6] While an exhaustive literature review is beyond the scope of this paper, we note that problems with this metric-based regulation approach have been well understood in the transplant literature (Chandraker et al. (2019)). In economics, Stith and Hirth (2016) examine the effects of the 2007 CMS implementation of Conditions of Participation quality standards based on 1-year post-transplant outcomes on TC strategic behavior. More kidney transplant candidates were removed from the waitlist for being "too sick" and fewer kidney transplants were undertaken following a TC's breaching of CMS's quality tolerance band relative to TCs that did not. However, little is known about whether and how TC post-transplant metric-based regulations and/or OPO regulations that impose a standard on donors per "eligible death referral" affect OPO recovery activities or strategic behaviors.

Our study is the first to evaluate the potential impact of regulations on OPO and TC behavior together, and provide evidence for the potential benefits of a holistic regulatory approach.

Kessler and Roth (2012) and Kessler and Roth (2014) study how changes in the

[6] Examples of this work include but are not limited to: Schold et al. (2010); Schold et al. (2013a); White et al. (2015); Schold et al. (2019); Schold et al. (2013b); Schold and Howard (2006); Buccini et al. (2014); Weinhandl et al. (2009); Howard et al. (2009); Abecassis et al. (2009); Massie and Segev (2013); Kasiske et al. (2019); Goldfarb (2020).

rules governing organ waiting lists might impact potential donors' decision to register as an organ donor.[7] Here, we propose a novel laboratory game design that can provide a foundation for future research on the market design and regulations surrounding the supply chain for transplantable organs. This paper follows the tradition of market design experiments that proved to be an important way for economists to communicate with the participants in and administrators of a market being studied. Studies by Kagel and Roth (2000), McKinney et al. (2005), Chen and Sönmez (2006), Bolton et al. (2013), Kagel et al. (2010), Kagel et al. (2014), Goeree and Lindsay (2020), and Budish and Kessler (2022) are all experiments on market design that had direct effects on the adoption and implementation of designs involving medical residents, gastroenterology fellows, school choice, eBay reputation mechanisms, bandwidth auctions, and course allocation. In the case of studying market design for organ transplantation, our lab experiment offers a simple platform to study the impact of different incentive architectures presented by current and counterfactual regulatory regimes.

The paper is organized as follows. Section 2 provides a brief overview of the policies and incentives facing the US deceased-donor organ transplantation industry. Section 3 describes the experiment and empirical strategy. It includes a replication of the original experiment with different treatment parameters, to separate the effect of differing parameters from the effect of aligning the incentives of TCs and OPOs. Section 4 describes the data and presents the key results. Section 5 discusses the implications of this study for transplantation policy, and Section 6 concludes.

2 Background: Organ Transplantation Policy Today

2.1 Transplant Center Regulations

TC regulation has gone through a few iterations. The Transplantation Amendment Act of 1990 led to the first publicly reported transplant center-specific report from United Network for Organ Sharing (UNOS) and OPTN in 1992 (reporting results from October 1, 1987 to December 31, 1989). In 1993, the OPTN/UNOS Board of Directors approved using center-specific reports to "flag" programs for poor performance but it was not until

[7]Those studies were followed by field studies of how such changes played out in Israel (Stoler et al. (2017), Stoler et al. (2016)). It also led to a subsequent literature (e.g., Herr and Normann (2016)).

1997 that OPTN MPSC published a method for "flagging" underperforming transplant programs (Kasiske et al. (2019), Jay and Schold (2017), Chandraker et al. (2019), Luskin and Nathan (2015)).

A major shift in TC incentives came when the 2007 CMS regulations for transplant programs were published in the Final Rules for Approval and Reapproval. Under these rules, a minimum standard was set for 1-year post-transplant patient and graft survival as conditions for center certification and maintenance of funding with CMS. At the same time, the OPTN MPSC published similar but not identical 1-year post-transplant metrics to "flag" TCs. Based on this, the OPTN can designate a TC as a member-not-in-good-standing and impose costly peer reviews while jeopardizing TCs' center of excellence designations with private insurers. A loss of center of excellence status can spell a loss of privately insured patients and a drop in reimbursements. In October 2019, CMS eliminated one-year post-transplant outcome requirements as a condition of Medicare recertification of TCs (Centers for Medicare and Medicaid Services (2019)) but OPTN continues to impose outcomes standards. As of the end of 2021, OPTN continues to analyze and publish detailed data on one-year post-transplant performance and can recommend that a TC be shut down for not meeting performance benchmarks (Phend (2020)).

The use of a single metric, one-year post-transplant graft and patient survival, for identifying under-performing transplant programs by various bodies is not without its critics. Reliance on a metric that focuses only on two outcomes of patients selected into transplantation but ignores patients who remain on the waitlist, strongly disincentivizes the use of imperfect organs. Although the MPSC recognizes the need to evaluate multiple phases of transplant care and not just look at post-transplant, short-term outcomes, reforms to this approach to performance management for TCs remained at discussion stages as of late 2021 (Procurement and Network (2021)).

2.2 Organ Procurement Organization Regulations

CMS evaluates and recertifies OPOs every 4 years. Recertification is based on compliance with two outcome measures that proceed from the donor conversion metric (how often do "eligible" deceased patients become potential donors) and the yield metric (how often are donated organs actually transplanted). The donor conversion and yield metrics in their current formulation were adopted in 2000.[8] While no OPO in the last 20 years has been decertified for poor performance, LiveOnNY was threatened with a shutdown in 2014 and then again in mid-2018 due to poor performance scores for nearly a decade and organ recovery rates that were among the lowest in the nation (Kindy and Bernstein (2019)). OPOs can also incur significant costs associated with review and quality improvement initiatives, which can both be triggered by bad performance metrics.

One aspect of these metrics that has been particularly controversial is the "eligible deaths" denominator of the main OPO metrics. "Eligible" referrals are determined to be medically suitable if the patient has been declared brain dead on a ventilator, and does not have any of a defined list of medical diagnoses that preclude organ transplantation. The number is reported by the OPOs, and some commentators (Goldberg et al. (2017), Rosenberg et al. (2020)) argue that this gives OPOs the room to cherry-pick deceased donors to recover, passing over potential donors whose organs are likely not acceptable to TCs. A 2011 OPTN review of OPOs reports "large inconsistencies and variations in how OPOs reported [eligible deaths] data" (Luskin and Nathan (2015)).

While the degree of manipulability of reported "eligible deaths" numbers remain controversial, it is clear that regulation of OPOs based on the two metrics in their current form can inadvertently offer incentives to under-recover organs. First, "eligible" does not include Donation after Circulatory Determination of Death, which makes up $15-20\%$ of actual donors today. It also does not include older donors over 75, who could potentially benefit *some* transplant candidates (Aubert et al. (2019)), especially older candidates or candidates who expect a longer wait time for a kidney. Second, the decisions to put a patient on a ventilator and to invest the neurology resource to declare brain death can be influenced by multiple parties, including the OPO. A risk averse decision maker might want to make decisions that lower the odds that a donor on the margin of being acceptable to TCs for transplantation is declared brain dead in time (before their heart stops).

More importantly, reliance on the donation rate and transplantation rate metrics to judge OPOs neglects a key driver of OPO behavior: TCs' propensity to accept and use an organ for transplantation. OPOs have little incentive to recover organs from a deceased

[8]The same 2000 legislation also established the donor service area monopolies and extended OPO recertification from two to four years to recognize that small OPOs had significant swings in donation potential year to year and to give time for performance improvement efforts to mature.

donor if those organs will not be accepted by any TC. Regulating OPOs without aligning TC's interests with increased organ recoveries creates fragmented incentives that may lead to inefficiencies.

On November 20, 2020, President Trump's Executive Order on Advancing American Kidney Health led to a new CMS final rule that updates OPO Conditions of Participation (Trump (2019), Lentine and Mannon (2020), Centers for Medicare and Medicaid Services (2020)). Beginning in 2022, the "eligible deaths" denominator will be replaced by the number of organ donors in the OPO's donor service area as a percentage of inpatient deaths among patients 75 years old or younger with a primary cause of death that is consistent with organ donation. This final rule will also impose stringent re-certification requirements including one that designates OPOs with below-median donation rate and transplantation rate measures as a Tier 3 OPO. The curent plan is that Tier 3 OPOs will be decertified and will not be able to compete for any other open donor service area contracts.

Whether these new OPO regulations will be constructive is unclear.[9] Donors making donations after circulatory determination of death and donors over 75 are still not included in the metrics' denominator and recovery efforts targeting them therefore not directly incentivized. The regulation of TCs remain not directly aligned with that of OPOs.

In summary, as of late 2021, TCs and OPOs face separate incentives. TCs face incentives to get good post-transplant health outcomes for candidates that they can select into receiving a kidney transplant, but no direct incentives to improve the health of those who remain untransplanted on their kidney waitlist. OPOs face incentives to have few recovered-but-not-transplanted organs, and hence not to recover all possible organs, especially organs from older donors who will not enter the denominators of the OPO performance metrics. OPOs might also limit the resources devoted to recovering from donors likely to yield organs of a lower quality that might not meet the acceptance thresholds of risk averse TCs. We explore the consequences of such incentives in our lab experiment.

[9]The executive branch, through the Department of Health and Human Services, has direct authority over the regulations overseeing the key players in the kidney transplantation supply chain. Furthermore, improving the regulation of organ procurement and transplantation has received bipartisan support. These suggest that the federal government can feasibly enact policies that resemble the holistic treatment that we investigate in this paper (especially if we are able to address the practical issues outlined in the Discussion).

3 Experimental Design

We model the decision environment facing OPOs and TCs with a 2-player multi-round game. In the game, urns containing either red or blue balls represent patients, jars containing red or blue balls represent kidneys, the mixing of the balls from a jar into an urn represents a kidney transplantation, blue (red) balls represent a good (bad) health outcome, and the drawing of a ball from an urn (mixed with a jar or not) represents the manifestation of the health outcome.

Each pair of subjects plays 10 rounds of the game. Upon arrival to the game with another subject, each subject is randomly assigned a role as either Player 1 or Player 2. At the beginning of each round, Player 1 will receive a pair of identical jars (which can be thought of as a pair of kidneys from a deceased donor but are presented to subjects only as jars of balls) and Player 2 will receive 2 non-identical urns (representing two different patients on the transplant waitlist). The jars and the urns each have 100 balls at the beginning of each round, each ball is either blue or red. The percentage of balls that are blue in an urn can be thought of as the baseline probability of survival or good health outcome by other measures for a patient. One of Player 2's urns has a lower expected number of blue balls than the other. This could be thought of as the situation where a TC has sicker and healthier patients on the waitlist.

In each round, Player 1 chooses whether to offer his pair of jars to Player 2.[10] The round ends if Player 1 chooses not to offer Player 2 the pair of jars. Player 1 can only offer two jars or none. If Player 1 offered the pair of jars to Player 2, Player 2 makes a decision to either decline the jars, mix all the balls from one of the two jars into only one of her urns, or mix all the balls from each of the two jars into each of her urns (one jar into each urn). Think of Player 1's decision as a rough representation of an OPO's decision to recover kidneys after getting a referral for a potential eligible deceased donor: offering a pair of jars can be interpreted as declaring a deceased person eligible for donation and recovering their two kidneys. Think of Player 2's decision as a representation of a TC's decision to accept one or both kidneys for transplantation or not. See Figure I for a schematic of the stage game played each round.

[10]Subjects in our experiment may include individuals who identify with different genders across the gender spectrum. For expository purposes, we will use he/him pronouns for player 1 (player representing the OPO), and use she/her pronouns for player 2 (player representing the TC).

Insert Figure I about here.

One ball is drawn from each one of Player 2's urns at the end of each round. Transplants can change the risk of a red ball, by mixing the jar (kidney) with the urn (patient). If Player 2 mixed the balls in a jar into one of her urns, the draw from that urn is made after the balls from the jar were mixed in (ie. the health outcome is assessed post-transplant). This is a highly simplified representation of the manifestation of patient outcomes: if a red ball is drawn from an urn, it is akin to a bad health outcome for the patient represented by that urn.

3.1 Information Available to Players

We are modeling the OPO's information at the point of deciding whether to recover kidneys from a deceased person, i.e. before the surgical recovery has taken place, and the TC's information after examining the patients and the recovered kidneys.

It is common knowledge to both players that there are 2 types of urns ("High Blue" and "Low Blue") and 2 types of jars ("Low Quality" and "High Quality"). Player 2 will have one urn (patient) of each type in every round while Player 1 will get a pair of identical jars.

In each round, Player 1 will not be able to observe the exact number of blue balls in each of the jars that he received but will be told whether he received a pair of "High Quality" or "Low Quality" jars for that round, each with probability 1/2. A "High Quality" jar has a number of blue balls (out of 100) drawn from a uniform distribution $U[70, 100]$ whereas a "Low Quality" jar has a number of blue balls (out of 100) drawn from a uniform distribution $U[0, 70]$. Player 1 is also aware that Player 2 has an urn of each type. He knows that a "High Blue" urn has a number of blue balls (out of 100) drawn from a uniform distribution $U[40, 100]$, and a "Low Blue" urn has a number of blue balls (out of 100) drawn from a uniform distribution $U[0, 60]$. While he knows that Player 2 will have an urn of each type, he will neither observe nor receive any signals about the exact number of blue balls in either of her urns.

Unlike Player 1, Player 2 can see the actual number of blue balls in the jars as well as the actual number of blue balls in each of her urns before making any decisions in each round. This is a simplified way to represent the knowledge that a TC has about the transplant candidates, and kidneys already recovered from a deceased donor and offered to the TC.[11]

It is common knowledge to both players that at the time of decision Player 1 can only observe jar and urn types while Player 2 can observe the exact number of blue and red balls in each jar and urn.

The actual instruction screens are in the online Appendix. The software for the game can be accessed through the Dataverse Repository of the Journal.

3.2 Control and Treatment Conditions

Pairs of subjects are randomly assigned at the beginning of the game to have either a payoff scheme that represents the status quo rules governing OPOs and TCs or a payoff scheme that represents our proposed holistic regulatory rules. The former rewards (penalizes) OPOs for TC-accepted (TC-declined) kidneys from deceased donors declared by OPOs to be eligible and rewards (penalizes) TCs for good (bad) patient health outcomes for transplants that TCs chose to carry out by accepting a kidney. In contrast, the proposed holistic approach emphasizes aligning the interests of OPOs and TCs by rewarding each of them for improvements in good health outcomes both among patients selected for a transplantation and patients who remain on the waitlist.

In the *status quo* condition, subjects are paid based on different payoff schemes. For each round, Player 1 earns \$0.10 by offering two jars that were subsequently both accepted by Player 2, gets a \$0.30 penalty by offering two jars that were subsequently both declined by Player 2, and gets \$0.00 by either not offering the jars at all or offering two jars when only one jar was accepted by Player 2.[12] For each round, Player 2 earns nothing if Player 1 did not offer the jars or if she declined both jars, earns \$0.25 by mixing a jar with an urn and drawing a blue ball from the mixed urn, and gets a \$1.00 penalty instead if a red ball was drawn from the mixed urn.[13] Under these payoffs where Player 2 is paid

[11]Information about recovered kidneys can include the report of the recovering surgeon, photos, biopsies, etc.

[12]By giving equal payoffs to the OPO for not recovering one or for recovering two and having one discarded (kidney declined by the TC), we remain agnostic about the relative costs of unrecovered versus discarded kidneys, both of which are subjects of considerable controversy (Recent Senate hearings (in 2022) have focused on discard kidneys, while Aubert et al. (2019) address the issue of unrecovered kidneys in the US as compared to France).

[13]In other words, Player 2 gets a \$0.50 reward for mixing jars into both urns and drawing blue balls from both, a \$0.75 penalty for mixing jars into both urns and drawing a blue ball from one and a red ball from the other, and a \$2.00 penalty for mixing jars into both urns and drawing red balls from both.

only if she mixed jars and urns, it is sometimes income-maximizing for Player 2 to mix a jar with fewer blue balls (lower quality) into an urn as long as the mixed urn still yields a high enough chance to draw a blue ball (notice that this "makes an urn worse" and can be thought of as an inappropriate transplant: the patient's expected health outcome would be better in the absence of a transplant.).[14]

In the *holistic* condition, both subjects are paid based on the number of blue and red balls drawn from *both* urns in each round. For each round, Player 1 earns \$0.16 for each blue ball drawn from the urns whether mixing happened or not. He gets an \$0.08 penalty for each red ball drawn from the urns. Player 2 earns \$0.20 for each blue ball drawn from the urns whether mixing happened or not. She gets a \$0.10 penalty for each red ball drawn from the urns.

Note that, while in the status quo condition only the health outcomes of transplanted patients are used for the determination of TC payoffs, in the holistic condition health outcomes of both patients determine payoffs every round.

In practice, if we want to make TCs responsible for an entire patient pool attributed to them, it would be unreasonable to penalize them at the same level for a death or a bad health outcome for every patient as is presently done for transplanted patients (since not all patients can be transplanted, and untransplanted patients will inevitably experience a substantial rate of bad health outcomes). The lower penalty for red balls drawn from urns at the end of each round of the holistic condition reflects this intuition.

That said, keep in mind that for this treatment of the experiment, we are not only changing from status quo to holistic incentives, but also changing the payoff parameters for TCs. We will return to this later, when we replicate both conditions with different parameters.

Subjects were told exactly how many rounds they would play the game, as well as the payoff scheme for both themselves and the other player with whom they were playing the 10 rounds of the game. At the end of each round, subjects were informed of the payoffs they received for that round as well as what the other player received for that round. Both players were presented with a table to help them keep track of actions taken by both players and their respective earnings from each of the previous and current rounds.

All experimental sessions were conducted online using US-based subjects recruited via

[14]For example, in an actual TC, a patient could sometimes be better served if the present organ were rejected and instead the patient waited for a better offer.

the Prolific platform. Batches of subjects were given a 10-minute window to show up and participate in the game. Subjects' basic demographic information were collected as they arrived, and then were admitted to a waiting room. They were then paired off based on arrival time to the waiting room to play 10 rounds of the game.

3.2.1 Theoretical Predictions

In this section, we describe the subgame perfect equilibrium under expected income maximization by both players as well as the optimal response by the first player given the actual behavioral response of the second player.

Insert Figure II about here.

First, consider the status quo condition. Examine Figure II as we work backwards from Player 2's optimal strategy. An expected income maximizing Player 2 will only mix a jar into the urn if the proportion of blue balls in the mixed urn is high enough. Let $\#(Blue)$ denote the number of blue balls in a mixed urn (with 100 balls from a jar and a 100 balls from an urn), each possible $\#(Blue)$ given a combination of urn and jar types is represented by a point on a graph in Figure II. As both urns and jars are drawn from uniform distributions, each point on a graph is equally likely to be drawn given the specific jar and urn types depicted by that graph. Player 2 would prefer mixings (transplants) represented by points to the Northeast on the graph. An expected income maximizing Player 2 will mix a jar into an urn if:

$$0.25\#(Blue) \geq \#(Red) = 200 - \#(Blue) \tag{1}$$

where $\#(Red)$ is the number of red balls (out of 200). Therefore, Player 2 will use a threshold rule of mixing if the mixed urn will have $\#(Blue) \geq 160$. The transplants that are accepted to an expected income maximizing Player 2 are represented by the (green) shaded areas in Figure II.

Reviewing the top panel of graphs in Figure II, we can see that an expected income maximizing Player 2 would leave jars unused in most cases (as represented by the small shaded areas for most of the graphs on the top panel). Reviewing the first two graphs on the top panel, we can see that Player 2 will mix a High Quality jar into a High Blue urn less than 42% of the time and into a Low Blue urn less than 1% of the time. Reviewing the last two graphs on the top panel, we can see that Player 2 will mix a Low Quality jar into a High Blue urn less than 2% of the time and never into a Low Blue urn.

Next we work backwards and evaluate what Player 1 would do under the status quo condition. Let $\Pr(Accepted = x)$ denote the probability that Player 2 will accept x of the 2 offered jars. Given Player 2's optimal strategy, an expected income maximizing Player 1 will only offer if:

$$0.1 \Pr(Accepted = 2) \geq 0.3 \Pr(Accepted = 0). \tag{2}$$

Given Player 2's optimal strategy, we have $\Pr(Accepted = 2) < 1\% < \Pr(Accepted = 0)$. This makes the expected income of Player 1 strictly negative whenever he recovers a pair of jars. Therefore, Player 1's optimal strategy is to never recover any jars.

If all players are maximizing expected income perfectly and it is common knowledge that they do, we would expect that very few transplants would be performed from recovered organs in the status quo condition and no organs would be recovered as a result.

The subgame perfect equilibrium looks different in the holistic condition. To see this, first consider Player 2 again. Player 2's optimal strategy is now represented by the middle panel of graphs in Figure II. Player 2 mixes a jar into an urn whenever it can improve the odds of drawing a blue ball from the resulting urn. That is, Player 2 would prefer mixings (transplants) represented by points Southeast of the line on which the number of blue balls is the same for the jar and the urn on the graph.

Reviewing the first two graphs on the middle panel in Figure II, we can see that Player 2 will mix a High Quality jar into a High Blue urn most of the time (75%) and into a Low Blue urn 100% of the time. Reviewing the last two graphs on the middle panel, we can see that Player 2 will mix a Low Quality jar into a High Blue urn about 11% of the time and into a Low Blue urn about 58% of the time. Player 2 will mix 1.2 jars per round on average (1.8 if the jars are High Quality and 0.7 if the jars are Low Quality).

Player 1 wants Player 2 to have the option to increase the chance of a blue ball being drawn at the end of each round. Given Player 2's optimal strategy, Player 1 can only weakly reduce his expected income by not offering the jars and offering the jars is a dominant strategy, regardless of jar quality.

We get the following for expected income maximizing players:

Proposition: In the subgame perfect equilibrium under the assumption that players maximize expected income, jars are never recovered by Player 1 (OPO) under the status quo condition and jars are always recovered under the holistic condition, Player 2 (TC) will perform no mixing (transplants) under the status quo condition but 1.2 mixes will be conducted per round under the holistic condition.

Neither actual transplant centers nor experimental subjects are robots who maximize expected income perfectly. An important deviation from expected income maximization is action bias. Based on results from our pilot experiments,[15] we see much higher rates of transplantation (mixing) than expected under the status quo condition, offering support for the presence of action bias for Player 2.[16]

Action bias describes people's tendency to favor action over inaction, sometimes at a cost. In our experiment action bias could result in part from other-regarding preferences, since both players can benefit only when the other acts: e.g. Player 2 might act to reciprocate being offered jars by player 1.

Such a natural bias for action has been documented in and out of the laboratory (Ledyard (1995), Patt and Zeckhauser (2000), Zeelenberg et al. (2002), Bar-Eli et al. (2007), Sunstein and Zeckhauser (2011)). It is reasonable to model hospitals that opt into having a transplant service as agents who exhibit action bias, just as others have in other healthcare provider settings (e.g., Kiderman et al. (2013)). This bias might have translated to risky actions (e.g. risky transplants) that led to bad outcomes and occasional penalties from UNOS, payers and other regulatory bodies. That is, the current regulations of TC's are designed to push back against this bias for action. Thus, using lab experiment participants who similarly exhibit action bias allows us to bring behavior absent from simple economic models to our experiment to model the players in transplantation.[17]

[15] Please see Appendix Section A.4 for more details including regressions analogous to the ones reported in this paper. The raw data and replication code for the pilots will be available for download on the Dataverse Repository of the Journal.

[16] While main results will be presented below, we offer a slight preview of the results to shed light on the degree of action bias. We observed in the status quo condition of the main experiment that Player 2s accepted both jars 42.0% and rejected both jars 5.0% of the time when the jar type is High Quality (thus, $0.1 \Pr(\mathit{Accepted} = 2) = 4.2\% > 0.3 \Pr(\mathit{Accepted} = 0) = 1.5\%$ for High Quality jars), and accepted both jars 16.8% and rejected both jars 39.2% of the time when the jar type is Low Quality (thus, $0.1 \Pr(\mathit{Accepted} = 2) = 1.7\% < 0.3 \Pr(\mathit{Accepted} = 0) = 11.7\%$ for High Quality jars). Considering Player 2's actual behavior, an expected income maximizing Player 1 will recover jars when the jars are High Quality and not recover them when they are Low Quality.

[17] There could be other reasons to explore these institutions behaviorally besides what is mentioned here directly. For instance, mistakes can also lead to actions that are different from what robots who maximize expected income would do.

The action bias for Player 2 can be large enough such that when the jars are High Quality the frequency with which Player 2 accepted both jars are more than one third of the frequency with which he accepted neither. Indeed, this is consistent with the data in the main experiment, presented below. Since $0.1\,\Pr(\mathit{Accepted} = 2) \geq 0.3\,\Pr(\mathit{Accepted} = 0)$ in this case, a Player 1 responding to a Player 2 who has such an action bias will recover High Quality jars.

If, we embrace bias for action for both Players to capture a behavioral trait of real-life OPOs and TCs, we get Player 2s who accept more jars for mixing than combinations that are represented by the green areas in Figure II and Player 1s who offers both High and Low jars under the status quo condition and not just under the holistic condition.

We get the following for expected income maximizing players in an environment with a bias for action of the magnitudes we saw in our pilot experiments and in this one:
Behavioral best response to players with action bias: In best response play for players with a bias to act, jars of High Quality are always recovered by Player 1 (OPO) under either conditions and jars of Low Quality are sometimes recovered under status quo condition but always recovered by Player 1 under holistic condition, and Player 2 (TC) will perform mixings (transplants) under either conditions.

3.2.2 Separating incentive architecture from payment magnitudes: a replication with different payoff parameters and additional controls

As noted earlier, the parameters for our Status quo and Holistic conditions were chosen to represent the current regulations and our proposed regulations, respectively. Status quo and holistic not only have different incentive structures but also have different payoffs for bad health outcomes, so our experimental results could be due to either or both. That is, we can entertain two hypotheses about different behaviors observed between the status quo and holistic treatments:

- *(our) main hypothesis*: Status quo discourages transplanting risky patients and kidneys, but holistic incentivizes every health-improving transplant; and
- *alternative hypothesis*: status quo severely penalizes failed transplants but (because of the different payoff parameters) holistic doesn't.

That is, the alternative hypothesis suggests that if we reduced the large penalty for failed transplants in the status quo condition, it might elicit behavior more like the holistic condition, even without directly providing holistic incentives. In the replication experiment described next, which employs quite different payoff parameters, we look at a new "counterfactual" status quo treatment that does not severely penalize failed transplants (but continues to only incentivize TC's on the success or failure of transplants that are performed). We will compare it to a re-parameterized holistic treatment that continues to incentivize both OPO and TC based on the outcomes of all patients.

The replication experimental design, information availability to the players, and control/treatment conditions are all identical to those of the main experiment, except for the payoff parameters, i.e. the benefits from drawing a blue ball and the costs of drawing a red ball for both conditions.

The replication payoff parameters are also chosen to equalize the payoffs at equilibrium for expected payoff maximizers, and the expected payoff (of 0) for taking no action (In the holistic treatment, players who take no action nevertheless receive payoffs based on the health outcomes of all patients.).[18]

We consider a "Counterfactual Status Quo" condition that doesn't severely penalize bad transplant outcomes. This allows us to compare it to a comparable "Holistic treatment 2" such that the expected payoffs from the two conditions are the same at equilibrium. The conditions described here were run as a replication and sensitivity analysis sixteen months after the main sample.

The new parameters are as follows (As will be explained below, some of the parameters have to be specified to multiple decimal places to equalize all of the payoffs described above.).

Counterfactual Status Quo: Player 1's payoffs are the same as in the Status Quo condition, Player 2's payoffs are: the penalty of drawing a red ball after mixing is \$0.1 and the reward of drawing a blue ball after mixing is \$0.11175. Here, we reversed the magnitude of the penalty of drawing a red ball after mixing and the reward of drawing a blue ball after mixing for Player 2 relative to the main experiment.

Holistic Treatment 2: Player 1 earns \$0.0957 for each blue ball drawn and loses

[18] There are no rewards or penalties when both players do nothing under the status quo condition while the odds of drawing blue balls and red balls are both 50% (as we make one draw from $U[40, 100]$ and another one from $U[0, 60]$). Therefore, to ensure that players who do nothing will get the same expected income (\$0) under either condition (Holistic Treatment 2 and Counterfactual Status Quo), we set the reward for drawing a blue ball under Holistic Treatment 2 to equal the penalty for drawing a red ball.

\$0.0957 for each red ball drawn from the urns whether mixing happened or not, while Player 2 earns \$0.1 for each blue ball drawn and loses \$0.1 for each blue ball drawn from the urns whether mixing happened or not.[19]

Notice that the Counterfactual Status Quo lowers the penalty but still penalizes only failed transplants, not bad outcomes for untransplanted patients. So, like our main status quo condition, it still discourages health improving transplants for very ill patients, but it allows more room for risky (and perhaps inappropriate) transplants of healthy patients, since it penalizes failures less. The Holistic Treatment 2 provides the same incentives as our original holistic treatment: every health-improving transplant is incentivized, but no inappropriate transplants.

Using similar theoretical reasoning as Section 3.2.1, we represent the optimal strategy for expected-income-maximizing Player 2 under the Counterfactual Status Quo condition in the bottom panel in Figure II. Reviewing the first two graphs on the bottom panel in Figure II, we can see that Player 2 is predicted to mix a High Quality jar into a High Blue urn 100% of the time and into a Low Blue urn 83% of the time. Reviewing the last two graphs on the bottom panel, we can see that Player 2 is predicted to mix a Low

[19] A few details regarding the reasoning behind the choice of these parameters, especially ones that carry many decimal places are explained here. Income maximizers under either holistic condition will improve the odds of drawing blue balls relative to the 50% chance by doing nothing whenever they can with a jar that has fewer blue balls than an urn. Recall that for High Quality jars blue balls are drawn from a $U[70, 100]$ distribution and for Low Quality jars blue balls are drawn from a $U[0, 70]$ distribution while an urn with number of blue balls drawn from a uniform distribution $U[40, 100]$ as well as another urn with number of blue balls drawn from a uniform distribution $U[0, 60]$ would be available in each round. If every opportunity to improve the odds of drawing blue balls were taken, Player 2's payoffs can be improved to improve the odds of drawing a blue ball to 61%. Picking a round number of \$0.1 for reward/penalty for Player 2 under Holistic Treatment 2, these improved odds of blue balls under optimal play will yield Player 2 an expected \$0.44 over 10 rounds. If we hold Player 1's payoff parameters constant between the original Status Quo condition and the new Counterfactual Status Quo condition while picking a round number of \$0.1 as the penalty of drawing a red ball after mixing for Player 2, we will have to set the reward of drawing a blue ball after mixing at \$0.11175 for Player 2 in the new Counterfactual Status Quo condition if we want Player 2 to earn \$0.44 over 10 rounds in expected value terms (as she would under Holistic Treatment 2) under the Counterfactual Status Quo condition when she acts (1) optimally as an expected income maximizer and (2) does so in response to a Player 1 who offers High Quality jars but not Low Quality jars. Subsequently, if a Player 2 acts in a manner consistent with (1) and (2) under these Counterfactual Status Quo condition payoff parameters and Player 1's payoff parameters are held constant between the original Status Quo condition and the new Counterfactual Status Quo condition, Player 1 will make an expected \$0.42 over 10 rounds. Therefore, we can pin down the payoff levels needed for Player 1 in the Holistic Treatment 2 under the improved odds of drawing a blue ball of 61% such that he would earn an expected \$0.42 over 10 rounds - giving us \$0.0957 for each blue ball drawn and loses \$0.0957 for each red ball drawn for Player 1 under Holistic Treatment 2. Simulations were conducted to identify and verify that the parameters up to the decimal places presented above will generate the same payoffs (accurate up to the 0.1 cent level per round) for an expected income maximizing player in a role (Player 1 or 2) regardless of which of two treatment conditions they are randomized into (see code posted on the Dataverse Repository of the Journal).

Quality jar into a High Blue urn about 64% of the time and into a Low Blue urn about 15% of the time. That is, an expected income maximizing Player 2 uses 1.83 of High Quality jars and 0.80 of Low Quality jars for mixing. In response to this optimal strategy, Player 1 will always recover/offer High Quality jars and never recover/offer Low Quality jars under the Counterfactual Status Quo condition. The optimal strategy for expected income maximizing Players 1 and 2 under Holistic Treatment 2 is unchanged from the original holistic treatment in the main experiment.[20]

Finally, we get the following for expected income maximizing players:

Proposition: In the subgame perfect equilibrium under income maximization, High Quality jars are recovered but Low Quality jars are not recovered by Player 1 (OPO) under the Counterfactual Status Quo condition and jars are always recovered under Holistic Treatment 2. Player 2 (TC) will perform 0.9 mixes on average per round under the Counterfactual Status Quo condition. Player 2 will perform 1.2 mixes on average under Holistic Treatment 2 (the same as under holistic treatment in the main experiment).[21]

4 Data and Results

4.1 Sample Population and Check for Balance

The experimental results for the two main conditions are from 324 subjects who participated in pairs in one of 162 sessions in the Summer of 2021.[22] There were 83 sessions under the holistic condition and 79 under the status quo condition. The experiment lasted up to 59:57 minutes and average earnings were $5.34 per subject, in addition to a $5.00 show-up fee. The experiment was conducted using software written in JavaScript and hosted on AWS cloud environment. Subjects were recruited through Prolific and redirected to the game website after completing a brief Qualtrics survey on their demographic background. Payoffs and actions are recorded in a PostgreSQL relational database that

[20] Player 2's optimal strategy under Holistic Treatment 2 is represented by the middle panel in Figure II.

[21] At equilibrium, there can be some "inappropriate transplants" conducted by Player 2, who will perform 0.08 mixes per round in which a jar with fewer blue balls than an urn is mixed into the urn (nevertheless producing a posiive expected payoff). As Low Quality jars will not be offered at equilibrium, these bad transplants/mixes will not be with of Low Quality jars.

[22] 164 sessions were completed but one subject participated in 2 different sessions. The results in the paper uses the sample where this subject and the 2 players who played against them were dropped. Including the results from these 2 dropped sessions does not change the results in this paper.

can be downloaded as csv files.[23]

Table I presents summary statistics, with the full sample in columns 1 and 2. The last column of Table I investigates the balance between the treatment groups. Overall, balance is achieved across subject demographics (with the exception of the age groups 35-44 and 45-54) as well as the number of blue balls in the jars and urns that players encounter.

Insert Table I about here.

Panel A of Figure III presents descriptive statistics that compares the key outcomes of the status quo and holistic conditions, and each of the subsequent sections will discuss the bar clusters shown in the Figure.

Insert Figure III about here.

4.2 Recovery, Discards, Transplants

The first pair of bars of Figure III displays the average rates of jar recovery and offering by Player 1. Overall, Player 1s recover 17.4 percentage points fewer jars under status quo compared to holistic regulations: 58.6% and 76.0% of the jars are recovered under status quo and holistic conditions respectively. The first column of Table II shows the results unconditional on jar quality type while the second and third column show results conditional on jar quality type.[24] The difference is significant for low quality (fewer blue balls on average) jars: 29.9 percentage points fewer jars under status quo. Both of these results are significant. However, there is no statistically significant[25] difference in recovery rates when the jars are high quality jars. Player 1s recover similarly when the observed jar quality type is high (number of blue balls), but the recovery behavior diverges when the observed jar quality type is low.

Insert Table II about here.

[23]The software, along with the data and replication code, can be accessed through the Dataverse Repository of the Journal.

[24]Player 1s, the OPO players, can only observe jar quality type when they make the recovery decision but not actual quality (number of blue balls) Therefore, we are only controlling for jar quality types.

[25]At conventional levels.

The second pair of bars of Figure III and columns 4 to 6 of Table II show the results on instances where a pair of jars could have benefited at least one urn but was not recovered. Under the status quo condition, Player 1 "missed recovery opportunities" 26.2% of the time while missed recovery is only 15.7% under the holistic condition. The differences are significantly higher overall and when jars are of a low quality type, by 10.5 percentage points and 16.6 percentage points respectively (columns 4 and 5 of Table II). This indicates that one out of four times, it would been better for the Player 1 (the OPO) to recover the jars ("kidneys") for the objective of increasing the expected number of blue balls drawn from urns (better population health outcomes) but he did not under status quo.

The third pair of bars of Figure III and the first column of Table III show that the average quality (average number of blue balls) of recovered and offered jars is higher under the status quo condition. The average recovered jar has 65.1 blue balls (out of 100) under the holistic condition while the average recovered jar has 71.0 blue balls. This is because of the cherry picking done by Player 1s as evidenced by the lower recovery rates above: average quality of recovered is similar if we only look at jars that are ex ante high quality type.

Insert Table III about here.

The fourth pair of bars of Figure III and columns 2 to 4 of Table III display the results on jar discards by Player 2 when Player 1 offered her a jar. Despite the lower average quality of offered jars under the holistic regulations condition, Player 2s are not any more likely to discard jars than in status quo (see columns 2 to 4 of Table III). The threshold for accepting a jar to carry out a transplant is significantly lower under the holistic regulations condition as indicated by 7.0 fewer blue balls in the average post-transplant urn (see column 3 of Supplementary Table A4). The player representing TCs accepts "Low Quality" jars 22.1 percentage points less often when offered jars under status quo (Supplementary Table A4 column 6). More generally, she is also accepting jars less aggressively for transplants that increase the percentage of blue balls in an urn after transplantation in the status quo condition - the average jar (kidney) accepted for such transplants has 4.8 more blue balls under status quo (Supplementary Table A4 column 5). Under the holistic condition, they are more aggressively transplanting patients who are sicker on the baseline (urns transplanted under status quo has 13 more blue balls before transplantation; see Supplementary Table A4). Furthermore, the fifth pair of bars of Figure III and columns 5 to 7 in Table III show that Player 2s (TC) discard jars that could have helped the blue-ball-odds of an urn in a level that is 23.8 percentage points more under status quo condition (49.8%) than under holistic regulations (26.0%).

The sixth pair of bars of Figure III and Table IV displays the results on mixing ("transplants") by Player 2. The first 3 columns in Table IV shows the results on mixing rate. On average, Player 2s mix 35.5% of the urns under status quo and 44.3% of the urns under holistic conditions.[26] In other words, Player 2s mix about 8.8 percentage points less under status quo compared to holistic regulations. These results are significant at conventional levels, controlling for jar/urn quality or not (see Table IV).

Insert Table IV about here.

Conditional on jars being recovered and offered by Player 1, Player 2 can also harm the odds of a good outcome for an urn by mixing a jar with fewer blue balls than that urn into the urn ("harmful transplants"). The seventh pair of bars of Figure III and columns 4 to 6 in Table IV show that Player 2s conduct many more harmful transplants under the status quo condition (22.9%) than under holistic regulations (8.3%). These results are significant at conventional levels, controlling for jar/urn quality or not (see Table IV).

4.3 Impact on "Health Outcomes"

Based on the number of blue and red balls in the urns at the end of each round,[27] the expected number of red balls drawn is 46.2% and 42.1% for status quo and holistic conditions respectively (last pair of bars of Figure III).[28] In other words, we expect that

[26]Conditional on the actual recoveries and offers by Player 1 observed in the data for each round of each game, a perfectly expected income maximizing Player 2 would have mixed 9% of the urns under status quo and 48% of the urns under holistic conditions. The bias for action is statistically significant in the status quo condition but not in the holistic condition: the percentage of urns mixed is statistically similar at conventional levels between the observed and perfectly-expected-income-maximizing-Player-2-level conditional on actual recoveries for holistic but not for status quo.

[27]After mixing if mixing occurred. That is, although the players are paid on draws from the urns, we analyze the composition of the urns at the end of each round rather than the realization of the draws (which are a noisy signal of the composition of the urns).

[28]Conditional on the actual recoveries and offers by Player 1 observed in the data for each round of each game, a perfectly expected income maximizing Player 2 will mix in a way such that 41.6% and 38.8% of the balls drawn are expected to be red under status quo and holistic conditions respectively. While the levels of bad outcome rates are higher in the data than if the Player 2s are perfectly expected income maximizers, the comparative statics results are similar.

a red ball or a bad outcome is 4.1 percentage points more likely to be drawn from an urn under the status quo condition compared to under holistic regulations (see column 1 of Table V). To interpret the magnitude of the effect, we can compare these expected bad outcomes figures with the best attainable expected outcomes given the actual proportions of red and blue balls drawn for each round and each pair of player. The percent differences between the actual expected bad outcome figures and the benchmark best attainable expected outcomes can be thought of as measures of "excess bad health outcome."[29] The best attainable expected bad outcome, given the actual red and blue proportions of the urns and jars used in the game, is 39%. The excess expected bad outcomes are therefore 18% and 7% for the status quo and holistic conditions respectively. This means that moving from the status quo to the holistic condition reduced the excess expected bad outcomes by more than half.

A Mann-Whitney U test was also conducted on 162 pairs of players to determine if status quo treatment lead to a difference in mean expected bad outcomes. The status quo group has 79 pairs and the holistic group has 83 pairs. Results showed that the mean expected bad outcomes are significantly different between the two groups ($z = -4.996$, $p = 0.000$) at a significance level of 0.01. These results confirm that the holistic treatment had a significantly positive impact on the expected health outcomes.

In the actual draw in our experiment, 46.2% and 42.8% of the balls drawn are red under status quo and holistic conditions (see Supplementary Table A1).[30] The largest gains in health outcomes under holistic regulations are expected to come from the Low Blue urns (representing sicker patients) (see columns 2 and 3 in Table V).

Insert Table V about here.

The difference in bad outcome rate is most notable among the urns representing the healthiest and sickest patients. Looking only at rounds where at least one urn has more than 90% blue balls or less than 10% blue balls, a red ball is expected to be 7.0 percentage points more likely to be drawn from an urn under the status quo condition compared to

[29]Ie. "Excess bad outcome" = ("Observed expected bad outcome" divided by "Best attainable expected bad outcome") minus one.

[30]The difference of 3.4 percentage points is significant at the 10% level. Since we observe the actual lotteries associated with the drawing of a ball from each urn, the observed outcome is a noisy signal of the actual expected health outcomes compared to the exact expected value conditional on the realized urns. Therefore, we do not have to rely on the actual draws to understand the impact on expected health outcome.

under holistic condition (see column 4 of Table V).[31] Bad outcome differences are expected to be only 2.6 percentage points more likely under the status quo condition compared to under holistic regulations in rounds where neither urns have between 10% to 90% blue balls (see column 7 of Table V).[32]

The results on missed recoveries, bad discards, and bad mixings (transplants) offer evidence for efficiency loss. The results on red ball rates tell us that holistic regulations can benefit the "well-being" of the overall population, and this benefit will mostly come from improvements from the "sickest" and "healthiest."

4.4 Results Under Alternative Parameters (from Section 3.2.2)

The results for the experiment with alternative parameters are from the 314 subjects who participated in pairs in one of the 157 sessions in December of 2022 via Prolific. Participants in the main experiment were barred from participating in the experiment with alternative parameters. There were 70 sessions under Holistic Treatment 2 and 87 under the Counterfactual Status Quo condition. Table VI presents summary statistics in columns 1 and 2 and the test for balance between the treatment groups in the last column. Overall, balance is achieved across subject demographics (with the exception of the age group 25-34) and number of blue balls in the jars and urns, similar to the main experimental sample.

The comparative statics results for most of the key outcomes of interest (jar recovery rate, missed beneficial recovery, discard rate conditional on jar and urn blue ball count, bad discards, transplant rate, bad transplants, and "health outcomes") for this experiment with alternative parameters are the same as those obtained from the main experiment.[33] We discuss the key results from the experiment with alternative parameters briefly below.

[31]In the actual draw in our experiment, a red ball is $5-7$ percentage points more likely to be drawn from an urn at rounds where at least one urn has more than 90% blue balls or less than 10% blue balls under the status quo condition compared to under holistic regulations (see Supplementary Table A2).

[32]In the actual draw, bad outcome differences are not significant (at conventional levels) in rounds where neither urns have between 10% to 90% blue balls. See Supplementary Table A3.

[33]Given that the alternative parameters penalize red balls less when a transplant/mixing happened, we would expect that the effects of the Status Quo treatment on reduced transplant rate will be less pronounced while the increased bad transplant rates will be more pronounced under the alternative parameters when compared to the main experiment. Indeed, the comparison between the point estimates for both of these outcomes for the experiment with alternative parameters and the main experiment are as expected (although the difference is not statistically significant).

Comparing Table VII with Table II from Section 4.2, we can see that in the alternative parameter experiment, Player 1's recovery behavior (recovery rate and missed beneficial recovery opportunities) is very similar to the behavior of the Player 1s from the main experiment. Furthermore, the coefficients we obtained from the alternative parameter experiment are statistically indistinguishable from those we obtained from the main experiment.

Insert Table VII about here.

Next, we compare Table VIII with Table III from Section 4.2. Here we still find higher cherry-picking by Player 1 in recovery under Counterfactual Status Quo as the average quality (average number of blue balls) of recovered and offered jars is higher (significant at the 10% level) under the Counterfactual Status Quo. The magnitude of cherry-picking by Player 1 is slightly lower in the alternative parameter experiment (though not statistically different from the coefficient from the main experiment reported in Table III). Given this slightly lower level of cherry-picking by Player 1 and lower penalties for drawing a red ball after mixing by Player 2, it is perhaps not surprising that we now observe a significant difference in discard rate where there discards are 5.5 percentage point higher under Holistic Treatment 2 when we don't condition on the actual number of blue balls in the jars or urns (see Column 2 of Table VIII). However, this difference might not be material as the Player 2s are not any more likely to discard jars under Holistic Treatment 2 than in Counterfactual Status Quo once we control for the number of blue balls in just the jars or both the jars and urns (see columns 3 to 4 of Table VIII). Furthermore, the coefficients we obtained from the alternative parameter experiment are statistically indistinguishable from those we obtained from the main experiment.

Insert Table VIII about here.

Comparing Table IX with Table IV from Section 4.2, we can immediately see in Column 1 of Table IX that Player 2s mix about 5.8 percentage points less under Counterfactual Status Quo compared to holistic regulations under Holistic Treatment 2 (compared to 8.8 percentage points in the main experiment) and conduct many more harmful transplants under the Counterfactual Status Quo condition (Column 4-6 of Table IX). Again, the coefficients we obtained from the alternative parameter experiment are statistically indistinguishable from those we obtained from the main experiment.

Insert Table IX about here.

Finally, a quick comparison between Table X with Table V from Section 4.3 shows that the Counterfactual Status Quo treatment has very similar effect on the bad outcome rate (see Columns 1-9 of Table X). We conducted a Mann-Whitney U test on 157 pairs of players to determine if the Counterfactual Status Quo treatment lead to a difference in mean expected bad outcomes (relative to Holistic Treatment 2). Just like the main experiment, the results show that the mean expected bad outcomes are significantly different between the two groups ($z = -4.231$, $p = 0.000$) at a significance level of 0.01. These results confirm the finding that the Counterfactual Status Quo treatment had a significantly negative impact on the expected bad outcomes in the alternative parameter experiment, just as the status quo condition did in the main experiment.

Insert Table X about here.

Tables VII, VIII, IX, and X (and the pilot experiments reported in the Appendix Section A.4) demonstrate that the results and key insights of our experiment are robust to varying parameters. This suggests that the behavior and outcomes are driven by the architecture of the incentives rather than the levels and ratios of the benefit from the blue ball to the cost of the red ball.

5 Further design issues: which organizations have responsibility for which patients?

For simplicity, we have been referring to the patients for which transplant centers have responsibility as those on their waiting list. But in current practice, centers choose which patients to enroll on their waiting lists, while our holistic incentives call for a non-manipulable way to establish which TC and which OPO have responsibility for which patients. This is an issue that has begun to be addressed in other areas of healthcare, by establishing various sorts of patient service areas.

For example, the 2019 Advancing American Kidney Health initiatives begin to address the fact that multiple players influence patient health (by financially incentivizing dialysis centers to refer more patients to transplantation). At the heart of the initiative are new payment models to be rolled out in select Hospital Referral Regions (HRRs).[34] In these new payment models,[35] reimbursement payment adjustments made to "managing" providers (including nephrologists working at or outside of TCs, as well as dialysis centers) will be based on rates of home dialysis utilization and rates of kidney and kidney-pancreas transplantation among patients who are attributed to these providers' management.[36] We see the holistic regulation of TCs and OPOs as a complement to the direction taken by the federal government in this 2019 initiative.

The implementable policy that reflects holistic regulations would include deciding which OPO and TC are in charge of which patients. For example, the "default patient service area" for a TC might be based on areas from which past referrals to that TC originated, while the patient service area of an OPO can be derived from its current donor service area (which is typically much larger than the TC's area). The default patient attribution would make OPO's and TC's share responsibility for the patients in the intersection of their service areas.[37]

Attribution logics with similar approaches have already been widely adopted in other parts of healthcare.[38] This level of implementation won't be easy, and is beyond the scope of this paper, but will be an important part of the market design.

But integrated decision-making is hampered by the fact that organ policy remains splintered between different federal agencies and UNOS.

[34]HRRs are often used in health policy to aggregate patients into a region for care delivery purposes.

[35]The four highlighted payment models are Kidney Care First Model, Graduated Comprehensive Kidney Care Contracting (CKCC) Model, Professional CKCC Model, and Global CKCC Model.

[36]Without improving the incentive issues outlined in the present paper, the increase in referrals for transplantation might simply lengthen kidney waitlists.

[37]Default service areas would have to be adjustable to take account of patients enrolled at TCs outside of their default area, as well as patients enrolled at multiple TCs.

[38]Patient attribution to providers belongs to a broader, on-going policy discussion in healthcare. For instance, patients are attributed to an accountable care organization (ACOs) based on their patterns of primary care use, and each ACO is held accountable for the cost of all services received by the patients attributed to them, even those received outside the ACO. ACOs are also often measured and partially paid based on health outcomes for patients attributed to them (McClellan et al. (2010)). Even after over a decade, ACO patient attribution remains imperfect and prone to strategic manipulation. Interviews with industry experts reveal that there are private sector consultants who help ACOs strategically construct the most profitable patient panels under the terms of their respective payment contracts. Future work to identify a feasible attribution logic will be pivotal to the implementation of a holistic regulatory approach for organ transplantation.

6 Summary and Conclusion

The shortage of transplantable organs is the cause of severe limits on life-saving transplants in high income countries.[39] To increase access to transplantation, market design is needed to support every part of the supply chain of organs. These include at least

1. Increasing deceased donor registrations.[40]

2. Increasing deceased donor recoveries by OPOs, and transplants by TCs (this paper).[41]

3. Increasing living donation.[42]

4. More efficiently coordinating deceased and living donation through mechanisms like deceased-donor-initiated kidney exchange chains (Melcher et al. (2016)).

Discussions and some policy actions are already underway.

This paper uses a laboratory experiment to model the decision architecture facing OPOs and TCs, which shows that status quo regulations incentivise OPOs to under-recover kidneys and incentivize TCs to cherry-pick kidneys and patients. In the experiment, TCs missed opportunities to transplant that could have improved the health of patients and (instead) discarded recovered organs, and conducted some transplants that were not beneficial. OPOs failed to recover some organs that could have benefited patients.

Holistic regulation that aligned OPO and TC interests by rewarding them based on the health outcomes of the entire patient pool led, in our experiment, to more organ recoveries by OPOs, higher utilization of organs by TCs to transplant sicker patients who otherwise would have been untransplanted,[43] and more appropriate transplantation by TCs.

[39] In low and middle income countries, financial obstacles and availability of medical resources are more often the limiting factors.

[40] Kessler and Roth (2012), Kessler and Roth (2014), Stoler et al. (2016), Stoler et al. (2017).

[41] Advances in preserving organs would also increase transplants, by keeping organs viable for longer, and hence accessible to more potential recipients (Giwa et al. (2017)).

[42] See Ashlagi and Roth (2021) for a survey of kidney exchange. We also need to lower the barriers to cross-border exchange through global kidney exchanges (Rees et al. (2017)) and global kidney chains (Nikzad et al. (2021)).

[43] With better coordination with OPOs, TCs can prepare to respond promptly to offers of risky organs on behalf of patients who are far down the waitlist (by keeping them effectively "active" with recent health checks, etc.) so that they can swiftly respond and receive kidneys that are declined by patients higher up on the waitlist without delays that could push the organ's cold time beyond the acceptable range.

These findings suggest that we need to move beyond current discussions on the revision of isolated metrics, and move the regulation conversation to consider the TCs and OPOs together. Our experiment illustrates the benefits of an alternative, holistic regulatory approach but does not outline the details of implementation and the precise way a patient pool should be identified and attributed to different TCs and OPOs. Future work should translate the holistic regulatory approach into actual legislation and rules. For instance, we will need to develop a non-manipulable measure of transplant-eligible patients (distinct from currently TC-managed waitlists) and a logic to attribute patients to specific TCs. Nonetheless, insights from this study should alert regulatory bodies to the importance of aligned, holistic incentives.

References

Abecassis, Michael Messod, R Burke, GB Klintmalm, Arthur J Matas, RM Merion, D Millman, K Olthoff, JP Roberts, and American Society of Transplant Surgeons, "American Society of Transplant Surgeons transplant center outcomes requirements—a threat to innovation," *American Journal of Transplantation*, 2009, *9* (6), 1279–1286.

Ashlagi, Itai and Alvin E Roth, "Kidney exchange: an operations perspective," *Management Science*, 2021, *67* (9), 5455–5478.

Aubert, Olivier, Peter P Reese, Benoit Audry, Yassine Bouatou, Marc Raynaud, Denis Viglietti, Christophe Legendre, Denis Glotz, Jean-Phillipe Empana, Xavier Jouven et al., "Disparities in acceptance of deceased donor kidneys between the United States and France and estimated effects of increased US acceptance," *JAMA internal medicine*, 2019, *179* (10), 1365–1374.

Bar-Eli, Michael, Ofer H Azar, Ilana Ritov, Yael Keidar-Levin, and Galit Schein, "Action bias among elite soccer goalkeepers: The case of penalty kicks," *Journal of economic psychology*, 2007, *28* (5), 606–621.

Bolton, Gary, Ben Greiner, and Axel Ockenfels, "Engineering trust: reciprocity in the production of reputation information," *Management science*, 2013, *59* (2), 265–285.

Buccini, LD, DL Segev, J Fung, C Miller, D Kelly, C Quintini, and JD Schold, "Association between liver transplant center performance evaluations and transplant volume," *American Journal of Transplantation*, 2014, *14* (9), 2097–2105.

Budish, Eric and Judd B Kessler, "Can market participants report their preferences accurately (enough)?," *Management Science*, 2022, *68* (2), 1107–1130.

Centers for Medicare and Medicaid Services, "Medicare and Medicaid Programs; Regulatory Provisions to Promote Program Efficiency, Transparency, and Burden Reduction; Fire Safety Requirements for Certain Dialysis," 09 2019.

_ , "Medicare and Medicaid Programs; Organ Procurement Organizations Conditions for Coverage: Revisions to the Outcome Measure Requirements for Organ Procurement Organizations," 12 2020.

Chandraker, Anil, Kenneth A Andreoni, Robert S Gaston, John Gill, Jayme E Locke, Amit K Mathur, Douglas J Norman, Rachel E Patzer, Abbas Rana, Lloyd E Ratner et al., "Time for reform in transplant program–specific reporting: AST/ASTS transplant metrics taskforce," *American Journal of Transplantation*, 2019, *19* (7), 1888–1895.

Chen, Yan and Tayfun Sönmez, "School choice: an experimental study," *Journal of Economic theory*, 2006, *127* (1), 202–231.

Federation of American Scientists, "FAS Organ Procurement Organization Innovation Cohort Shares Data to Advance Organ Recovery Research," 10 2021.

Giwa, Sebastian, Jedediah K Lewis, Luis Alvarez, Robert Langer, Alvin E Roth, George M Church, James F Markmann, David H Sachs, Anil Chandraker, Jason A Wertheim et al., "The promise of organ and tissue preservation to transform medicine," *Nature biotechnology*, 2017, *35* (6), 530–542.

Goeree, Jacob K and Luke Lindsay, "The exposure problem and market design," *The Review of Economic Studies*, 2020, *87* (5), 2230–2255.

Goldberg, D, MJ Kallan, L Fu, M Ciccarone, J Ramirez, P Rosenberg, J Arnold, G Segal, KP Moritsugu, H Nathan et al., "Changing metrics of organ procurement organization performance in order to increase organ donation rates in the United States," *American Journal of Transplantation*, 2017, *17* (12), 3183–3192.

Goldfarb, David A, "Re: Time for Reform in Transplant Program-Specific Reporting: AST/ASTS Transplant Metrics Taskforce," *The Journal of urology*, 2020, *203* (1), 20–21.

Health Care Financing Administration, "Medicare and Medicaid programs: hospital conditions of participation; identification of potential organ, tissue and eye donors and transplant hospitals' provision of transplant related data," *Federal Register*, 1998, *63* (119), 33856.

Held, Philip J, Jennifer L Bragg-Gresham, Thomas G Peters, Frank McCormick, Glenn Chertow, William P Vaughan, and John P Roberts, "Cost structures of US organ procurement organizations," *Transplantation*, 2021, *105* (12), 2612–2619.

Herr, Annika and Hans-Theo Normann, "Organ donation in the lab: Preferences and votes on the priority rule," *Journal of Economic Behavior & Organization*, 2016, *131*, 139–149.

Howard, David H, "Why do transplant surgeons turn down organs?: A model of the accept/reject decision," *Journal of health economics*, 2002, *21* (6), 957–969.

Howard, Richard J, Danielle L Cornell, and Jesse D Schold, "CMS oversight, OPOs and transplant centers and the law of unintended consequences," *Clinical transplantation*, 2009, *23* (6), 778–783.

Husain, S Ali, Kristen L King, Stephen Pastan, Rachel E Patzer, David J Cohen, Jai Radhakrishnan, and Sumit Mohan, "Association between declined offers of deceased donor kidney allograft and outcomes in kidney transplant candidates," *JAMA network open*, 2019, *2* (8), e1910312–e1910312.

Jay, Colleen and Jesse D Schold, "Measuring transplant center performance: the goals are not controversial but the methods and consequences can be," *Current transplantation reports*, 2017, *4* (1), 52–58.

Kagel, John H and Alvin E Roth, "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment," *The Quarterly Journal of Economics*, 2000, *115* (1), 201–235.

_ **, Yuanchuan Lien, and Paul Milgrom**, "Ascending prices and package bidding: A theoretical and experimental analysis," *American Economic Journal: Microeconomics*, 2010, *2* (3), 160–85.

_ **,** _ **, and** _ , "Ascending prices and package bidding: Further experimental analysis," *Games and economic behavior*, 2014, *85*, 210–231.

Kasiske, Bertram L, Nicholas Salkowski, Andrew Wey, Ajay K Israni, and Jon J Snyder, "Scientific Registry of Transplant Recipients program-specific reports: where we have been and where we are going," *Current opinion in organ transplantation*, 2019, *24* (1), 58–63.

Kessler, Judd B and Alvin E Roth, "Organ allocation policy and the decision to donate," *American Economic Review*, 2012, *102* (5), 2018–47.

_ **and** _ , "Loopholes undermine donation: An experiment motivated by an organ donation priority loophole in Israel," *Journal of Public Economics*, 2014, *114*, 19–28.

Kiderman, Alexander, Itzhak Gur, Tali Bdolah-Abram, and Mayer Brezis, "in primary care: Evidence of action bias," *The Journal of Family Practice*, 2013, *62* (8).

Kindy, K and L Bernstein, "Despite low performance, organ collection group gets new federal contract," *Washington Post*, 2019.

Ledyard, John O, "Public Goods: A Survey of Experimental Research," in "The Handbook of Experimental Economics," Princeton University Press, 1995, pp. 111–194.

Lentine, Krista L and Roslyn B Mannon, "The Advancing American Kidney Health (AAKH) executive order: promise and caveats for expanding access to kidney transplantation," *Kidney360*, 2020, *1* (6), 557.

Luskin, R and H Nathan, "Eligible death statistic: not a true measure of OPO performance nor the potential to increase transplantation," 2015.

Massie, Allan B and DL Segev, "Rates of false flagging due to statistical artifact in CMS evaluations of transplant programs: Results of a stochastic simulation," *American Journal of Transplantation*, 2013, *13* (8), 2044–2051.

McClellan, Mark, Aaron N McKethan, Julie L Lewis, Joachim Roski, and Elliott S Fisher, "A national strategy to put accountable care into practice," *Health Affairs*, 2010, *29* (5), 982–990.

McKinney, C Nicholas, Muriel Niederle, and Alvin E Roth, "The collapse of a medical labor clearinghouse (and why such failures are rare)," *American Economic Review*, 2005, *95* (3), 878–889.

Melcher, Marc L, John P Roberts, Alan B Leichtman, Alvin E Roth, and Michael A Rees, "Utilization of deceased donor kidneys to initiate living donor chains," *American journal of transplantation*, 2016, *16* (5), 1367–1370.

Mone, Thomas D, "The business of organ procurement," *Current Opinion in Organ Transplantation*, 2002, *7* (1), 60–64.

National Academies of Sciences, Engineering, and Medicine, "Opportunities for organ donor intervention research: Saving lives by improving the quality and quantity of organs for transplantation," 2018.

_ , "A Fairer and More Equitable, Cost-Effective, and Transparent System of Donor Organ Procurement, Allocation, and Distribution - A Virtual Workshop," 2021.

Nikzad, Afshin, Mohammad Akbarpour, Michael A Rees, and Alvin E Roth, "Global kidney chains," *Proceedings of the National Academy of Sciences*, 2021, *118* (36), e2106652118.

Patt, Anthony and Richard Zeckhauser, "Action bias and environmental decisions," *Journal of Risk and Uncertainty*, 2000, *21*, 45–72.

Phend, Crystal, "Easing Transplant Outcome Scrutiny: A Small but Welcome Step," 01 2020.

Procurement, Organ and Transplantation Network, "Enhance Transplant Program Performance Monitoring System OPTN Membership and Professional Standards Committee," 2021.

Rees, Michael A, Ty B Dunn, Christian S Kuhr, Christopher L Marsh, Jeffrey Rogers, Susan E Rees, Alejandra Cicero, Laurie J Reece, Alvin E Roth, Obi Ekwenna et al., "Kidney exchange to overcome financial barriers to kidney transplantation," *American Journal of Transplantation*, 2017, *17* (3), 782–790.

Rosenberg, Paul, Michael Ciccarone, Bradley Seeman, David Washer, Gayle Martin, Jackson Tse, and Erica Lezama, "Transforming Organ Donation in America," *Bridgespan. org*, 2020.

Schold, JD and RJ Howard, "Prediction models assessing transplant center performance: can a little knowledge be a dangerous thing?," *American Journal of Transplantation*, 2006, *6* (2), 245–246.

_ , LD Buccini, ELG Heaphy, DA Goldfarb, AR Sehgal, J Fung, ED Poggio, and MW Kattan, "The prognostic value of kidney transplant center report cards," *American Journal of Transplantation*, 2013, *13* (7), 1703–1712.

Schold, Jesse D, Charlotte J Arrington, and Greg Levine, "Significant alterations in reported clinical practice associated with increased oversight of organ transplant center performance," *Progress in Transplantation*, 2010, *20* (3), 279–287.

_ , LD Buccini, TR Srinivas, RT Srinivas, ED Poggio, SM Flechner, C Soria, DL Segev, J Fung, and DA Goldfarb, "The association of center performance evaluations and kidney transplant volume in the United States," *American Journal of Transplantation*, 2013, *13* (1), 67–75.

_ , Rachel E Patzer, Timothy L Pruett, and Sumit Mohan, "Quality metrics in kidney transplantation: current landscape, trials and tribulations, lessons learned, and a call for reform," *American Journal of Kidney Diseases*, 2019, *74* (3), 382–389.

Siminoff, Laura A, Robert M Arnold, and Joan Hewlett, "The process of organ donation and its effect on consent," *Clinical Transplantation*, 2001, *15* (1), 39–47.

Solomon, Daniel A, Nicole Rabidou, Sanjay Kulkarni, Richard Formica, and Liana Fraenkel, "Accepting a donor kidney: an evaluation of patients' and transplant surgeons' priorities," *Clinical transplantation*, 2011, *25* (5), 786–793.

Stith, Sarah S and Richard A Hirth, "The effect of performance standards on health care provider behavior: evidence from kidney transplantation," *Journal of Economics & Management Strategy*, 2016, *25* (4), 789–825.

Stoler, Avraham, Judd B Kessler, Tamar Ashkenazi, Alvin E Roth, and Jacob Lavee, "Incentivizing authorization for deceased organ donation with organ allocation priority: the first 5 years," *American Journal of Transplantation*, 2016, *16* (9), 2639–2645.

_ , _ , _ , _ , and _ , "Incentivizing organ donor registrations with organ allocation priority," *Health Economics*, 2017, *26* (4), 500–510.

Sunstein, Cass R and Richard Zeckhauser, "Overreaction to fearsome risks," *Environmental and Resource Economics*, 2011, *48*, 435–449.

Trump, DJ, "Executive order on advancing American kidney health," *The White House https://www. whitehouse. gov/presidential-actions/executive-order-advancing-american-kidney-health*, 2019.

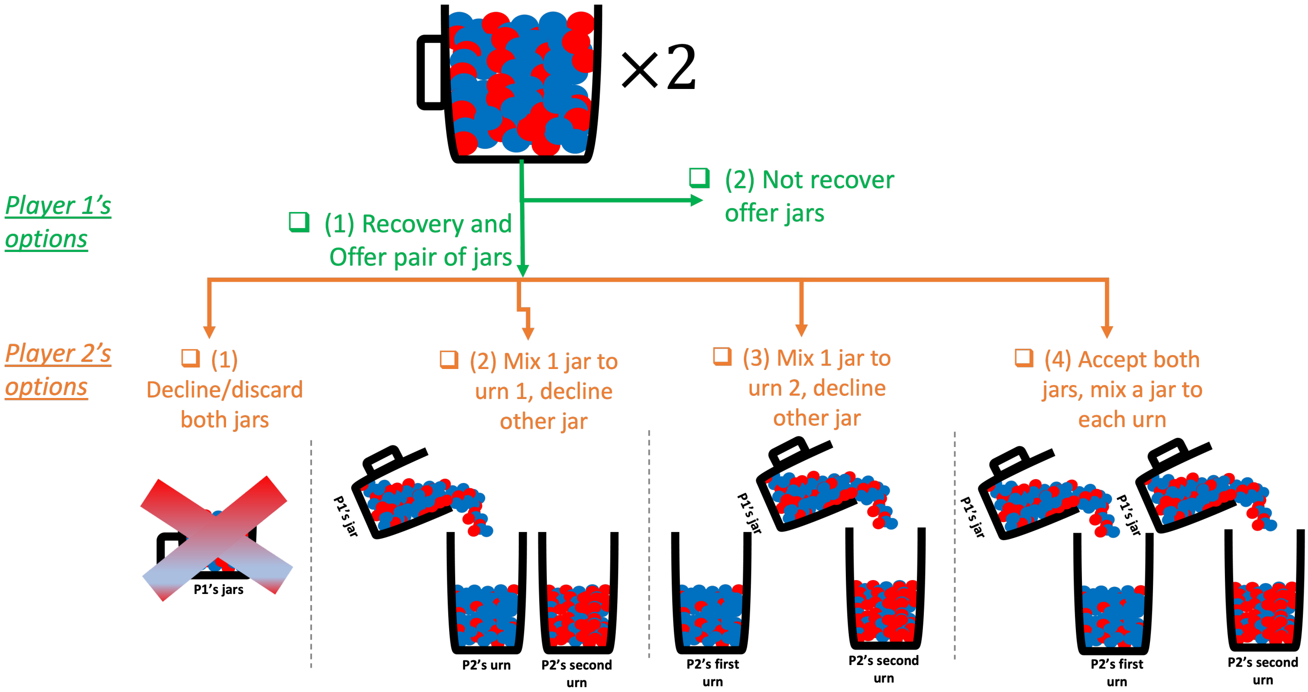
US Congress, "The Urgent Need to Reform the Organ Transplantation System to Secure More Organs for Waiting, Ailing, and Dying Patients," 05 2021.

Weinhandl, Eric D, JJ Snyder, AK Israni, and Bert L Kasiske, "Effect of comorbidity adjustment on CMS criteria for kidney transplant center performance," *American Journal of Transplantation*, 2009, *9* (3), 506–516.

White, Sarah L, Dawn M Zinsser, Matthew Paul, Gregory N Levine, Tempie Shearon, Valarie B Ashby, John C Magee, Yi Li, and Alan B Leichtman, "Patient selection and volume in the era surrounding implementation of Medicare conditions of participation for transplant programs," *Health services research*, 2015, *50* (2), 330–350.

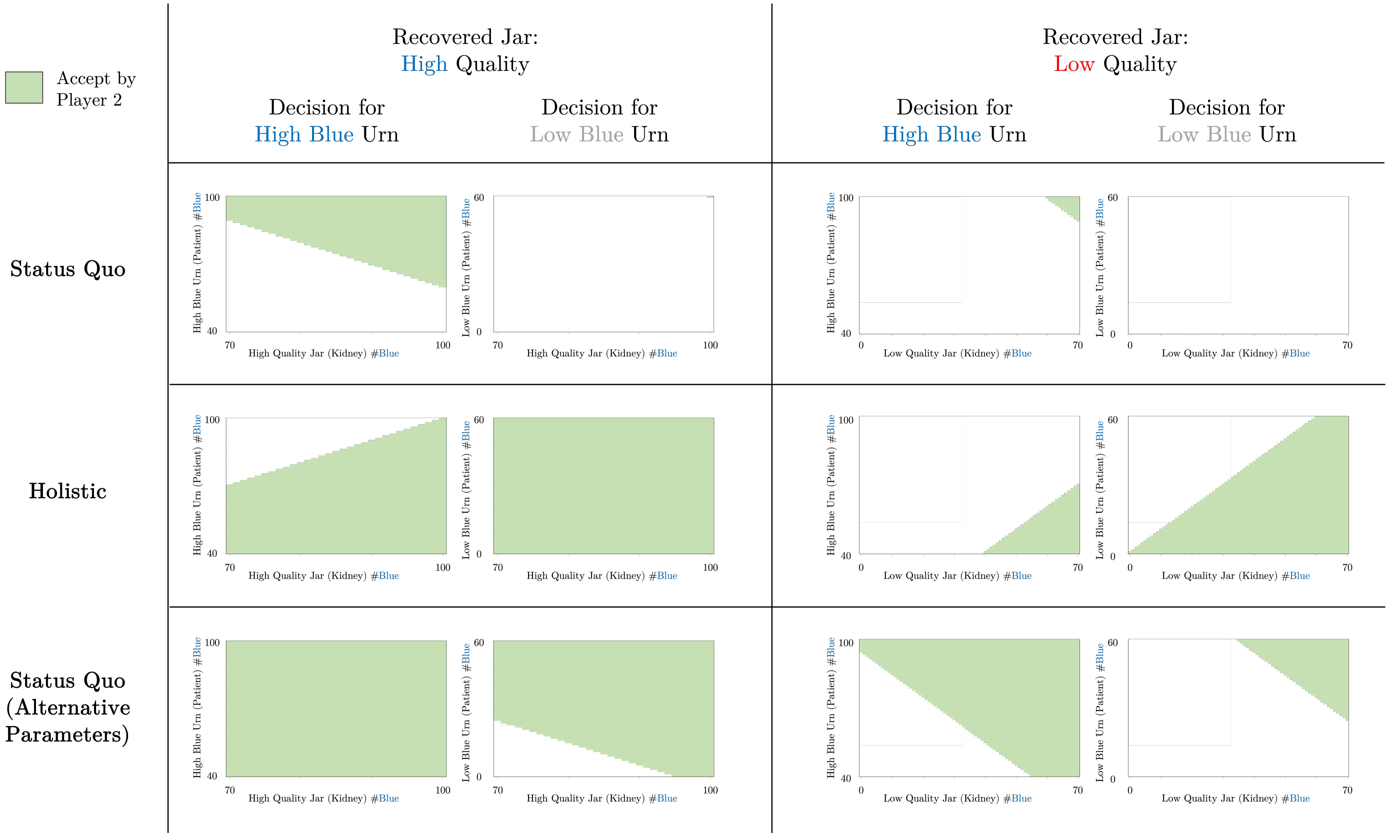
Zeelenberg, Marcel, Kees Van den Bos, Eric Van Dijk, and Rik Pieters, "The inaction effect in the psychology of regret.," *Journal of personality and social psychology*, 2002, *82* (3), 314.

Figure I: Stage Game for the OPO Player (P1) and TC Player (P2)



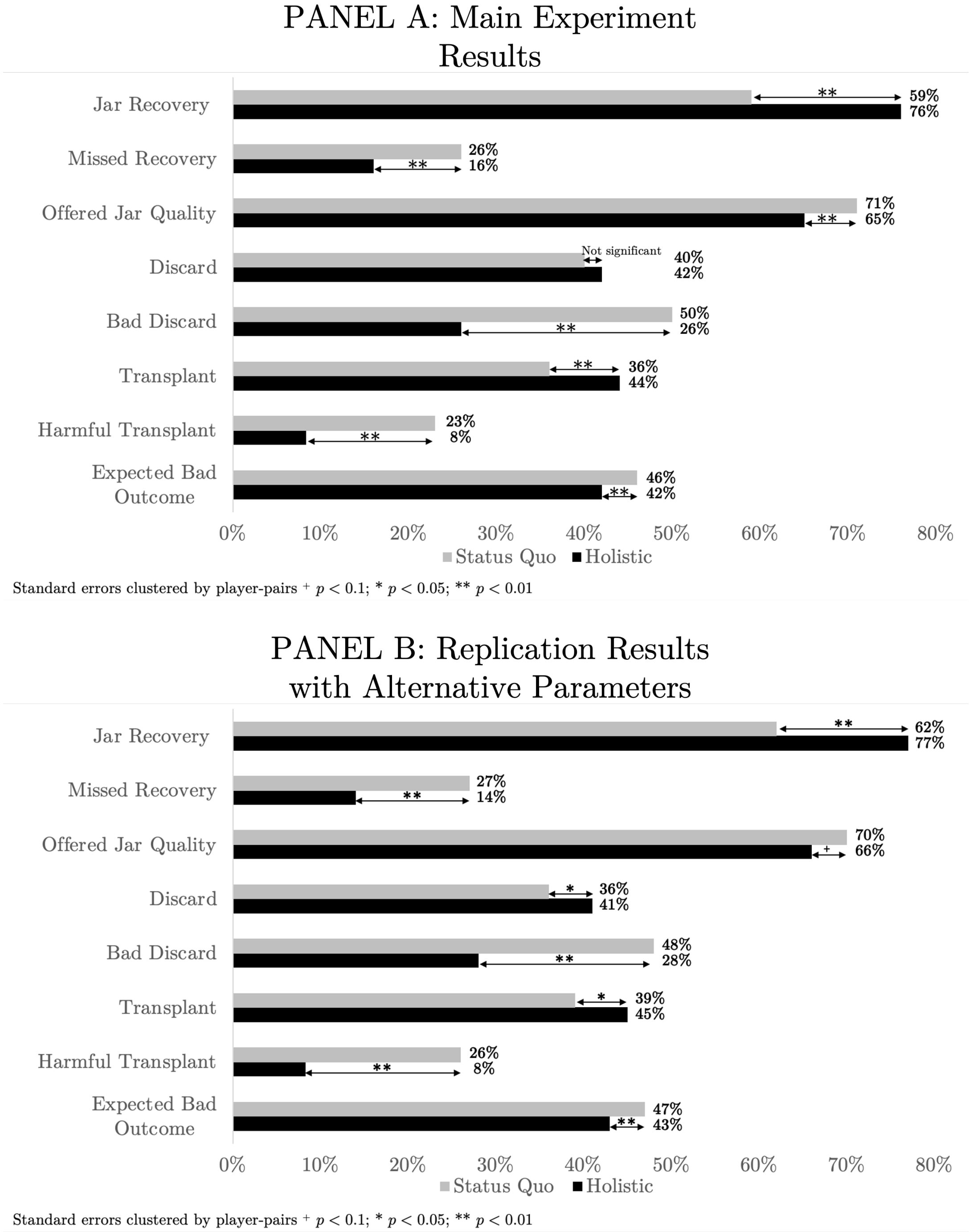
Notes: This figure presents the flow of the stage game in this experiment. Each pair of players play this stage game 10 times. The jars (with the handle) can be thought of as kidneys that can be recovered and offered. The urns can be thought of as transplant candidates. The chance of drawing a blue ball from an urn can be thought of as the odds of the patient represented by that urn getting a good health outcome.

Figure II: Jar Acceptance Strategy for Player 2 (Transplant Center) for Various Manifested Jars and Urns



Notes: This figure illustrates Player 2's (TC's) optimal strategy as an expected income maximizer. An expected income maximizing Player 2 will only mix a jar into the urn if the proportion of blue balls in the mixed urn is high enough. Each grid in each (of the 12) graph above represent the number of blue balls (out of 200) that would arise with each possible combination of urn and jar given the respective types. The green shaded area represent the jar-urn combination that would yield an expected income Player 2 non-negative expected income: the top panel illustrate optimal Player 2 strategy under status quo treatment under the baseline experimental parameters, the middle panel illustrate this for both the holistic treatment under the baseline experimental parameters as well as Holistic Treatment 2 under the alternative parameters (optimal strategy same), and the bottom panel illustrate Player 2's optimal strategy under the Counterfactual Status Quo treatment under the alternative parameters.

Figure III: Outcome Variable Comparison



Notes: This figure reports the average levels of the outcome variables of interest, by treatment condition, of the 324 subjects in the main experiment sample in Panel A and of the 314 subjects in the replication experiment (with alternative parameters) sample in Panel B. This Figure shows averages ("mean") by treatment condition. The significance-levels of the differences between treatment conditions are reported. "Jar Recovery" is the percentage of the pairs of jars that are recovered and offered by Player 1. "Missed Recovery" is the percentage of the pairs of jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1. "Offered Jar Quality" is the percentage of balls that are blue in each of the recovered and offered jars, "Discard" is the percentage of recovered jars that are discarded/rejected. "Bad Discard" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn). "Transplant" is the percentage of urns where a mixing happened. "Harmful Transplant" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn. "Expected Bad Outcome" is the proportion of red balls in an urn at the end of a round.

Table I: Summary Statistics and Balance Checks

	Total		Holistic		Status Quo		
	mean	s.d.	mean	s.d.	mean	s.d.	Difference p-value
Female	0.45	0.50	0.49	0.50	0.41	0.49	0.17
White	0.78	0.41	0.77	0.43	0.80	0.40	0.48
Black	0.09	0.28	0.08	0.27	0.09	0.29	0.60
Asian	0.07	0.26	0.09	0.29	0.05	0.22	0.16
Other Race	0.06	0.24	0.07	0.25	0.06	0.23	0.73
Hispanic	0.06	0.24	0.05	0.21	0.07	0.26	0.41
Employed Full-time	0.55	0.50	0.53	0.50	0.57	0.50	0.48
Unemployed	0.07	0.25	0.07	0.26	0.06	0.23	0.58
College Graduate	0.60	0.49	0.62	0.49	0.58	0.50	0.42
High School Graduate	0.99	0.08	1.00	0.00	0.99	0.11	0.15
Age 18-24	0.04	0.19	0.03	0.17	0.04	0.21	0.50
Age 25-34	0.33	0.47	0.36	0.48	0.29	0.46	0.18
Age 35-44	0.29	0.45	0.25	0.43	0.34	0.47	0.08*
Age 45-54	0.18	0.38	0.22	0.41	0.14	0.35	0.07*
Age 55-64	0.11	0.32	0.11	0.31	0.12	0.33	0.74
Age 65-74	0.04	0.20	0.02	0.15	0.06	0.23	0.13
Age 75-84	0.01	0.11	0.01	0.11	0.01	0.11	0.96
# Blue Balls in Jars	59.00	30.24	59.88	31.00	57.32	29.56	0.59
# Blue Balls in Urns	51.00	11.81	51.28	11.00	51.09	12.61	0.92
# Subjects	324		166		158		
N	3,240		1,660		1,580		

Notes: This table reports the background characteristics of the 324 subjects in the main sample, pooled and by treatment group. "Female" indicates the share of female sex; "White," "Black," "Asian," "Hispanic," and "Other Race" indicate the shares of subjects belonging to each of these categories. Age data was recorded in intervals, each one of the age categories indicate shares of subjects in these age buckets. "High School Graduate" indicate the share of subjects who did not graduate from high school, "College Graduate" indicate indicate the share of subjects who reported that they have a bachelor's or advanced degree. "Employed" indicate the share who are employed full-time, while "Unemployed" indicates the share of subjects who selected are not employed either full-time or part-time. "# Blue Balls in Jars," "# Blue Balls in High Blue Urn," and "# Blue Balls in Low Blue Urn" indicate the average number of blue balls (out of 100 balls that are either blue or red) in the jars and urns encountered by the subjects during the game. Table shows averages ("mean") and standard deviations ("s.d."). The Difference p-value column reports the p-value for the test of equality between the treatment and control groups. Stars indicate whether this difference is significant.

Table II: Impact on Jar Recovery Rates and Missed Opportunities for Beneficial Recoveries

	(1)	(2)	(3)	(4)	(5)	(6)
	Jar Recovery Rate			Missed Beneficial Recovery		
	All	Low Quality	High Quality	All	Low Quality	High Quality
Status Quo	-0.174**	-0.299**	-0.042	0.105**	0.166**	0.042
	(0.035)	(0.055)	(0.038)	(0.026)	(0.036)	(0.038)
Constant	0.760**	0.610**	0.913**	0.157**	0.225**	0.0874**
	(0.024)	(0.041)	(0.022)	(0.017)	(0.024)	(0.022)
N	1620	820	800	1620	820	800
R^2	0.035	0.090	0.004	0.017	0.032	0.004

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + \epsilon_i$. Player 1 (OPO) can either recover a pair of jars or not (1=recover/offer jar; 0=not recover/offer jar). The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The 1^{st} and 4^{th} columns reports results unconditional on jar quality type, the 2^{nd} and 5^{th} columns reports results conditional on the jar quality being low quality, and the 3^{rd} and 6^{th} columns reports results conditional on the jar quality being high quality. "Jar Recovery Rate" is the percentage of the pairs of jars that are recovered and offered by Player 1. "Missed Beneficial Recovery" is the percentage of the pairs of jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1.

Table III: Impact on Recovered Jar Quality, Discard Rates, and Discards that could have Benefited Urn(s)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Jar # Blue | Recovered	Jar Discard Rate | Recovered			% of Bad Discards		
Status Quo	5.985**	-0.022	0.020	0.024	0.238**	0.143**	0.060*
	(2.106)	(0.029)	(0.027)	(0.027)	(0.046)	(0.039)	(0.030)
Jar # Blue Balls			-0.007**	-0.007**		0.008**	0.010**
			(0.000)	(0.000)		(0.001)	(0.000)
High Urn # Blue Balls				0.003**			-0.013**
				(0.001)			(0.001)
Low Urn # Blue Balls				0.000			-0.002*
				(0.001)			(0.001)
Constant	65.060**	0.418**	0.878**	0.686**	0.260**	-0.191**	0.777**
	(1.379)	(0.021)	(0.030)	(0.054)	(0.032)	(0.021)	(0.053)
N	1094	1094	1094	1094	702	702	702
R^2	0.012	0.001	0.300	0.317	0.064	0.323	0.564

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either discard 0, 1 or 2 jars *if a pair of jars were offered by Player 1.* The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Jar # Blue | Recovered" is the number of blue balls in each of the recovered and offered jars (out of 100), "Jar Discard Rate | Recovered" is the percentage of recovered jars that are discarded/rejected. "% of Bad Discards" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn).

Table IV: Impact on Mixing ("Transplant") Rate, and Mixings that Gave an Urn Worse Odds for Blue

	(1)	(2)	(3)	(4)	(5)	(6)
	Mixing (Transplant) Rate			% Mixings Made Urn Worse		
Status Quo	-0.088**	-0.087**	-0.088**	0.146**	0.166**	0.163**
	(0.028)	(0.025)	(0.025)	(0.030)	(0.023)	(0.023)
Jar # Blue Balls		0.008**	0.008**		-0.008**	-0.008**
		(0.000)	(0.000)		(0.001)	(0.001)
High Urn # Blue Balls			-0.001**			0.004**
			(0.001)			(0.001)
Low Urn # Blue Balls			-0.000			0.002**
			(0.000)			(0.001)
Constant	0.443**	-0.018	0.086^{+}	0.083**	0.685**	0.331**
	(0.020)	(0.023)	(0.047)	(0.016)	(0.056)	(0.062)
N	1620	1620	1620	903	903	903
R^2	0.012	0.332	0.336	0.042	0.297	0.352

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either mix 0, 1 or 2 jars into the urns *if a pair of jars were offered by Player 1.* The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Mixing (Transplant) Rate" is the percentage of urns where a mixing happened, this is the number of transplant(s) divided by 2 per round. "% Mixings Made Urn Worse" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn.

Table V: Impact on Expected Bad Outcomes (Red Balls Drawn) Based on Actual Mixing Behavior

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All Urns		At Least One Urn Sickest or Healthiest			No Urn Sickest or Healthiest		
	All	Low Blue	High Blue	All	Low Blue	High Blue	All	Low Blue	High Blue
Status Quo	0.041**	0.055**	0.027**	0.070**	0.101**	0.039*	0.026**	0.034**	0.018*
	(0.007)	(0.012)	(0.008)	(0.016)	(0.022)	(0.015)	(0.007)	(0.011)	(0.009)
Constant	0.421**	0.578**	0.264**	0.401**	0.638**	0.164**	0.432**	0.548**	0.315**
	(0.006)	(0.008)	(0.006)	(0.011)	(0.016)	(0.010)	(0.006)	(0.008)	(0.007)
N	3240	1620	1620	1078	539	539	2162	1081	1081
R^2	0.007	0.020	0.007	0.012	0.054	0.015	0.004	0.010	0.004

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + \epsilon_i$. We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage of balls that are red balls in urns at the end of the round, representing the percentage of transplant candidates/patients expected to get a bad health outcome (e.g. expected mortality rate). The results here are conditional on actual Player 1 and Player 2 behavior in the game but not the actual draws. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The dependent variable in this table is the percentage of balls in an urn that is red at the end of a round.

Table VI: Summary Statistics and Balance Checks for Alternative Sample

	Total		Holistic		Status Quo		
	mean	s.d.	mean	s.d.	mean	s.d.	Difference p-value
Female	0.44	0.50	0.45	0.50	0.43	0.50	0.66
White	0.75	0.43	0.76	0.43	0.75	0.44	0.73
Black	0.09	0.29	0.10	0.30	0.09	0.28	0.68
Asian	0.08	0.27	0.08	0.27	0.08	0.27	0.95
Other Race	0.07	0.26	0.06	0.23	0.09	0.28	0.33
Hispanic	0.08	0.27	0.05	0.22	0.10	0.30	0.11
Employed Full-time	0.56	0.50	0.60	0.49	0.53	0.50	0.25
Unemployed	0.07	0.25	0.06	0.23	0.07	0.26	0.54
College Graduate	0.59	0.49	0.56	0.50	0.60	0.49	0.49
High School Graduate	0.99	0.08	0.99	0.09	0.99	0.08	0.88
Age 18-24	0.03	0.18	0.01	0.12	0.05	0.21	0.11
Age 25-34	0.34	0.47	0.40	0.49	0.29	0.45	0.04**
Age 35-44	0.30	0.46	0.26	0.44	0.33	0.47	0.18
Age 45-54	0.18	0.38	0.19	0.39	0.17	0.38	0.76
Age 55-64	0.11	0.31	0.09	0.28	0.13	0.33	0.25
Age 65-74	0.04	0.18	0.04	0.20	0.03	0.17	0.50
Age 75-84	0.01	0.11	0.01	0.12	0.01	0.11	0.83
# Blue Balls in Jars	62.22	28.22	59.19	29.26	64.67	27.19	0.23
# Blue Balls in Urns	48.63	11.96	47.00	11.29	49.95	12.35	0.13
# Subjects	314		140		174		
N	3,140		1,400		1,740		

Notes: This table reports the background characteristics of the 314 subjects in the alternative sample with revised experimental parameters, pooled and by treatment group. "Female" indicates the share of female sex; "White," "Black," "Asian," "Hispanic," and "Other Race" indicate the shares of subjects belonging to each of these categories. Age data was recorded in intervals, each one of the age categories indicate shares of subjects in these age buckets. "High School Graduate" indicate the share of subjects who did not graduate from high school, "College Graduate" indicate indicate the share of subjects who reported that they have a bachelor's or advanced degree. "Employed" indicate the share who are employed full-time, while "Unemployed" indicates the share of subjects who selected are not employed either full-time or part-time. "# Blue Balls in Jars," "# Blue Balls in High Blue Urn," and "# Blue Balls in Low Blue Urn" indicate the average number of blue balls (out of 100 balls that are either blue or red) in the jars and urns encountered by the subjects during the game. Table shows averages ("mean") and standard deviations ("s.d."). The Difference p-value column reports the p-value for the test of equality between the treatment and control groups. Stars indicate whether this difference is significant.

Table VII: Impact on Jar Recovery Rates and Missed Opportunities for Beneficial Recoveries for Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Jar Recovery Rate			Missed Beneficial Recovery		
	All	Low Quality	High Quality	All	Low Quality	High Quality
Status Quo	-0.156**	-0.253**	-0.055	0.132**	0.209**	0.055
	(0.029)	(0.056)	(0.035)	(0.024)	(0.039)	(0.035)
Constant	0.771**	0.632**	0.901**	0.140**	0.184**	0.099**
	(0.021)	(0.041)	(0.023)	(0.016)	(0.026)	(0.023)
N	1570	767	803	1570	767	803
R^2	0.028	0.063	0.007	0.026	0.051	0.007

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + \epsilon_i$. Player 1 (OPO) can either recover a pair of jars or not (1=recover/offer jar; 0=not recover/offer jar). The data here is from the alternative sample with revised experimental parameters. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The 1^{st} and 4^{th} columns reports results unconditional on jar quality type, the 2^{nd} and 5^{th} columns reports results conditional on the jar quality being low quality, and the 3^{rd} and 6^{th} columns reports results conditional on the jar quality being high quality. "Jar Recovery Rate" is the percentage of the pairs of jars that are recovered and offered by Player 1. "Missed Beneficial Recovery" is the percentage of the pairs of jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1.

Table VIII: Impact on Recovered Jar Quality, Discard Rates, and Discards that could have Benefited Urn(s) with Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Jar # Blue | Recovered	Jar Discard Rate | Recovered			% of Bad Discards		
Status Quo	3.953^{+}	-0.0550*	-0.0340	-0.0335	0.195**	0.120**	0.066*
	(2.101)	(0.0267)	(0.025)	(0.025)	(0.045)	(0.033)	(0.028)
Jar # Blue Balls			-0.005**	-0.005**		0.009**	0.010**
			(0.001)	(0.001)		(0.000)	(0.000)
High Urn # Blue Balls				0.002**			-0.012**
				(0.001)			(0.001)
Low Urn # Blue Balls				-0.000			-0.000
				(0.001)			(0.001)
Constant	65.740**	0.414**	0.763**	0.633**	0.283**	-0.252**	0.602**
	(1.465)	(0.019)	(0.035)	(0.058)	(0.034)	(0.022)	(0.059)
N	1075	1075	1075	1075	676	676	676
R^2	0.005	0.007	0.187	0.197	0.043	0.371	0.572

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either discard 0, 1 or 2 jars *if a pair of jars were offered by Player 1.* The data here is from the alternative sample with revised experimental parameters. The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Jar # Blue | Recovered" is the number of blue balls in each of the recovered and offered jars (out of 100), "Jar Discard Rate | Recovered" is the percentage of recovered jars that are discarded/rejected. "% of Bad Discards" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn).

Table IX: Impact on Mixing ("Transplant") Rate, and Mixings that Gave an Urn Worse Odds for Blue from Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Mixing (Transplant) Rate			% Mixings Made Urn Worse		
Status Quo	-0.058*	-0.056*	-0.057*	0.181**	0.186**	0.176**
	(0.026)	(0.025)	(0.025)	(0.031)	(0.025)	(0.024)
Jar # Blue Balls		0.001**	0.006**		-0.008**	-0.008**
		(0.000)	(0.000)		(0.001)	(0.001)
High Urn # Blue Balls			-0.001+			0.005**
			(0.001)			(0.001)
Low Urn # Blue Balls			0.000			0.002**
			(0.001)			(0.001)
Constant	0.452**	0.0458	0.103+	0.083**	0.638**	0.237**
	(0.019)	(0.029)	(0.053)	(0.017)	(0.051)	(0.065)
N	1570	1570	1570	920	920	920
R^2	0.005	0.233	0.235	0.056	0.306	0.367

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either mix 0, 1 or 2 jars into the urns *if a pair of jars were offered by Player 1.* The data here is from the alternative sample with revised experimental parameters. The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Mixing (Transplant) Rate" is the percentage of urns where a mixing happened, this is the number of transplant(s) divided by 2 per round. "% Mixings Made Urn Worse" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn.

Table X: Impact on Expected Bad Outcomes (Red Balls Drawn) Based on Actual Mixing Behavior from Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All Urns			At Least One Urn Sickest or Healthiest			No Urn Sickest or Healthiest		
	All	Low Blue	High Blue	All	Low Blue	High Blue	All	Low Blue	High Blue
Status Quo	0.040**	0.052**	0.029**	0.055**	0.077**	0.032*	0.033**	0.039**	0.026**
	(0.009)	(0.012)	(0.009)	(0.015)	(0.020)	(0.015)	(0.009)	(0.012)	(0.009)
Constant	0.426**	0.581**	0.270**	0.413**	0.659**	0.168**	0.432**	0.539**	0.325**
	(0.006)	(0.009)	(0.006)	(0.011)	(0.014)	(0.011)	(0.006)	(0.008)	(0.006)
N	3140	1570	1570	1080	540	540	2060	1030	1030
R^2	0.007	0.017	0.007	0.007	0.033	0.009	0.007	0.013	0.008

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + \epsilon_i$. The data here is from the alternative sample with revised experimental parameters. We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage of balls that are red balls in urns at the end of the round, representing the percentage of transplant candidates/patients expected to get a bad health outcome (e.g. expected mortality rate). The results here are conditional on actual Player 1 and Player 2 behavior in the game but not the actual draws. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The dependent variable in this table is the percentage of balls in an urn that is red at the end of a round.

A Online Appendix

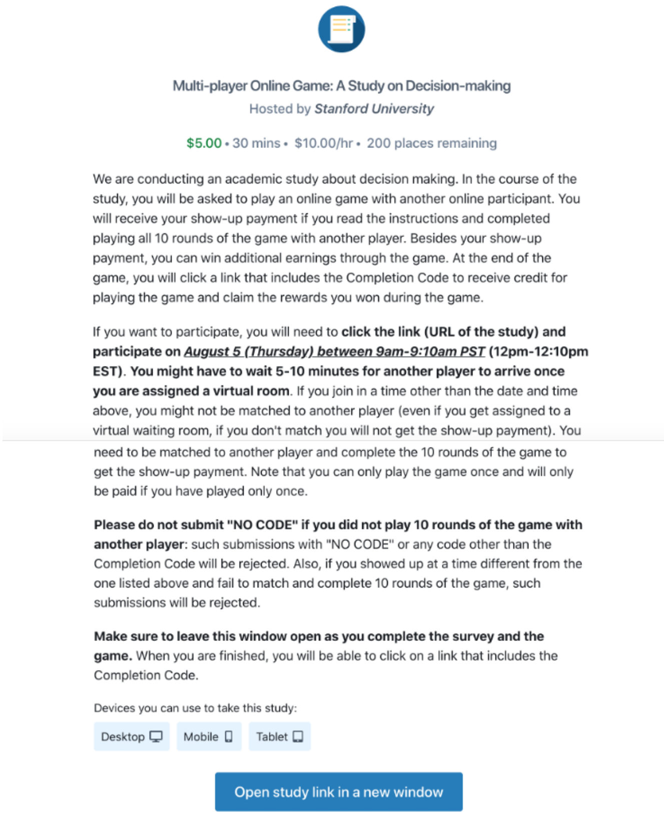
This is the online appendix for "Regulation of Organ Transplantation and Procurement: A Market Design Lab Experiment" by Alex Chan and Alvin E. Roth.

A.1 Experimental Details

In this Section, we describe the design of our experiment in detail, including the recruitment screen on Prolific (A.1.1), and the game screens (A.1.2).

A.1.1 Recruitment On Prolific

Subjects were recruited on the Prolific platform (see Section 4.1). The subjects in the game participated between August 2 2021 and August 12 2021. Prolific recruitment posting were posted during 7 days (August 2, 3, 4, 5, 7, 9, 12) in this period where subjects were recruited to play the game during a 10-minute window on each of these 7 days. The narrow window for participation is to increase the number of subjects who are online simultaneously, so that we can pair them off into our two-player game. See Figure A1 for the recruitment information on Prolific. Subjects who managed to get paired off become part of our sample.


Multi-player Online Game: A Study on Decision-making
Hosted by *Stanford University*

$5.00 • 30 mins • $10.00/hr • 200 places remaining

We are conducting an academic study about decision making. In the course of the study, you will be asked to play an online game with another online participant. You will receive your show-up payment if you read the instructions and completed playing all 10 rounds of the game with another player. Besides your show-up payment, you can win additional earnings through the game. At the end of the game, you will click a link that includes the Completion Code to receive credit for playing the game and claim the rewards you won during the game.

If you want to participate, you will need to **click the link (URL of the study) and participate on *August 5 (Thursday) between 9am-9:10am PST* (12pm-12:10pm EST). You might have to wait 5-10 minutes for another player to arrive once you are assigned a virtual room**. If you join in a time other than the date and time above, you might not be matched to another player (even if you get assigned to a virtual waiting room, if you don't match you will not get the show-up payment). You need to be matched to another player and complete the 10 rounds of the game to get the show-up payment. Note that you can only play the game once and will only be paid if you have played only once.

Please do not submit "NO CODE" if you did not play 10 rounds of the game with another player: such submissions with "NO CODE" or any code other than the Completion Code will be rejected. Also, if you showed up at a time different from the one listed above and fail to match and complete 10 rounds of the game, such submissions will be rejected.

Make sure to leave this window open as you complete the survey and the game. When you are finished, you will be able to click on a link that includes the Completion Code.

Devices you can use to take this study:

Desktop Mobile Tablet

Open study link in a new window

Figure A1: Recruitment screen posted on Prolific

A.1.2 Overview of Experiment

After a participation consent screen, subjects demographic information were collected (see Section A.1.6). After these short steps, the subjects are re-directed to a game page where they will enter their Prolific ID (see Figure A2) and get assigned to a online waiting room until they are matched with another subject (see Figure A3). Pairing is done randomly based on arrival time to the game. Each pair is then randomized into a treatment condition (status quo or holistic) using Bernoulli draws as subjects arrive.

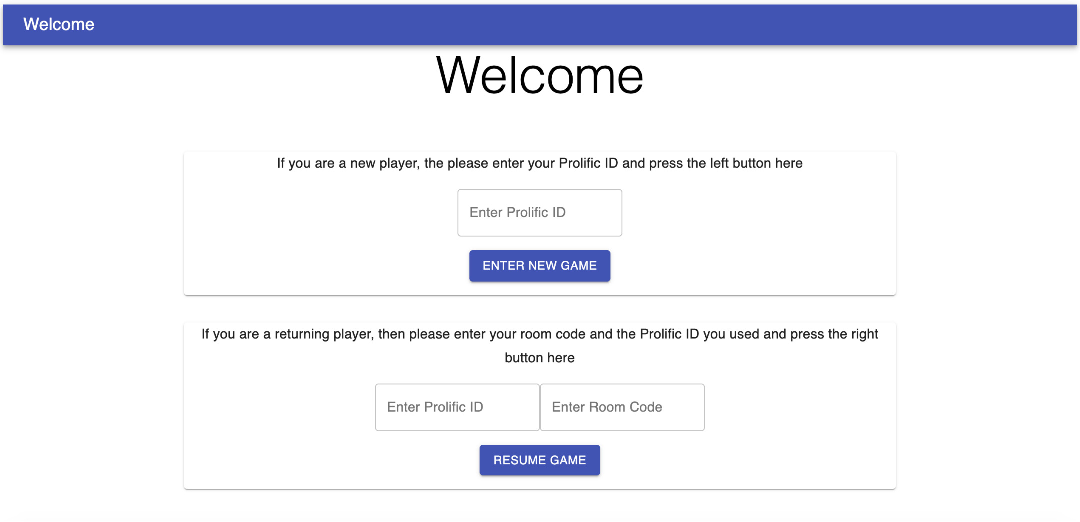


Figure A2: Game Launch Page

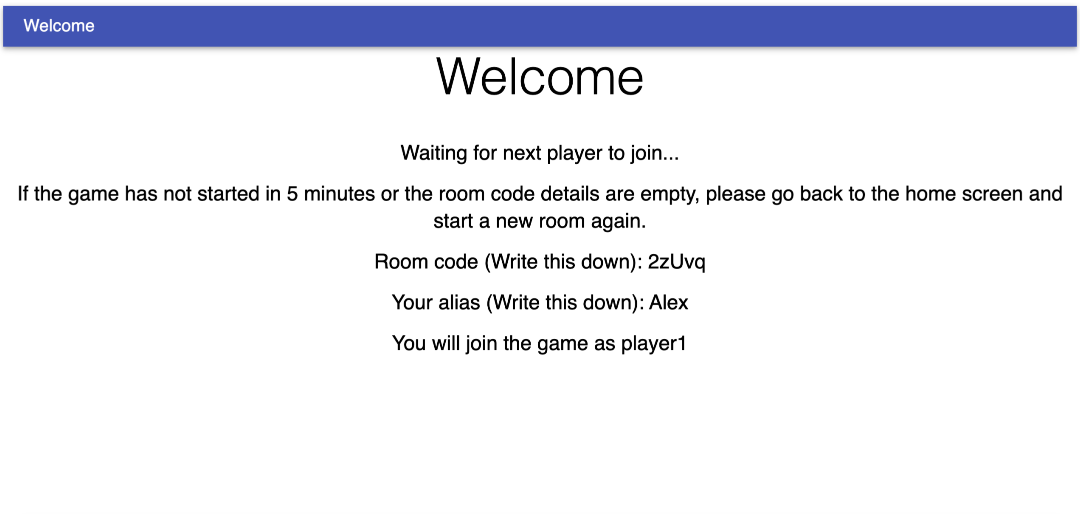


Figure A3: Game Online Waiting Room Page

This Appendix Section should serve to provide additional details about the actual experience of subjects of this experiment as they play this online, two-player game. We will share the screens broken down into the following three sections (Sections A.1.3, A.1.4 and A.1.5).

A.1.3 Game Introduction

Subjects randomized into either the status quo and holistic conditions will go through 6 introductory screens that are identical. These 6 screens serve to explain game details like the available actions to both players (Figures A4 and A9), number of rounds (Figure A4), high-level description of the games including the fact that balls will be drawn from urns at the end of each round (Figure A5), the nature of the jars and urns (how many, the distribution from which balls are drawn from) (Figures A6, A7), as well as information available to each player (Figure A8). Essentially, the first 6 screen outlines the game as described in Section 3 except information regarding the actual payoff schemes.

(Page 1 of 10)

Welcome to this study of decision-making.

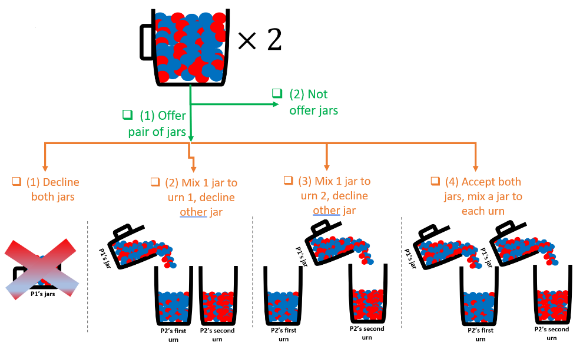
The study will take about 30 minutes including the instructions. The instructions are simple, and if you follow them carefully, you can earn tokens in addition to your $5.00 show-up payment. You will receive your show-up payment if you and the player paired with you finish all 10 rounds of the game. Besides your show-up payment, you will be given a "Bank" of 500 tokens (equivalent to $5.00) to which additional earnings will be added and from which penalties will be deducted. Tokens will be translated into dollars at the end of the game (1 token = $0.01). If your Bank's balance is negative, you will receive only your show-up payment. The show-up payment and your final "Bank" balance will be paid to you after the study ends.

You will be randomly assigned a role as either Player 1 or Player 2. You will play 10 rounds of a game with another participant.

At the beginning of each round, Player 1 will receive a pair of identical jars and Player 2 will receive 2 non-identical urns. The jars and the urns each have 100 balls at the beginning of each round, each ball is either blue or red.

Each round of the game goes roughly as follows (see figure on the right):

- Player 1 chooses whether to offer his/her pair of jars to Player 2. The round ends if Player 1 chooses not to offer Player 2 the pair of jars.
- If Player 1 offered the pair of jars to Player 2, Player 2 makes a decision to either (1) decline the jars or (2)/(3) mix all the balls from ONE of the two jars into ONLY ONE of his/her urns, or (4) mix all the balls from EACH of the two jars into EACH of his/her urns



Wait 5 seconds before the next page can appear...

‹ 1 2 ›

Figure A4: Game Screen 1

(Page 2 of 10)

The computer takes one draw from each one of Player 2's urns at the end of each round. A total of 2 balls are drawn each round. After each draw, the drawn ball's color will be recorded and the ball replaced into the urn from which it was drawn. The draws will be made at the end of the round. If Player 2 mixed the balls in a jar into one of his/her urns, the draw from that urn is made after the balls from the jar were mixed in.

Your earnings will depend on various factors including your decisions, the decisions of the other Player, and how many blue balls were drawn from Player 2's urns at the end of the round.

Wait 5 seconds before the next page can appear...

‹ 1 2 3 ›

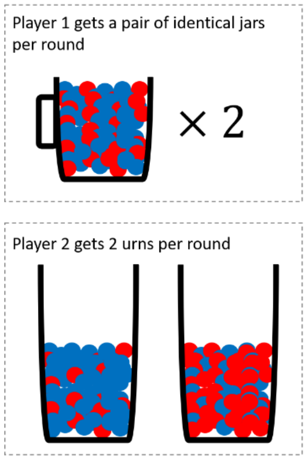
Figure A5: Game Screen 2

(Page 3 of 10)

Instructions for Player 2

In this study, you have been assigned the role of Player 2. You have been randomly matched with another participant who will be in the role of Player 1. Your earnings will depend on your decisions, as well as on the decisions of Player 1. There will be 10 rounds of this study. For all 10 rounds you will be paired with the same Player 1, who will participate at the same time as you.

In each of the 10 rounds Player 1 will be given two identical jars of balls and you will be given two urns of balls. A different pair of jars and 2 different urns will given for each round (the jars or urns do not carryover to the next round(s)). Player 1's earnings for each round can depend on whether Player 1 offered the pair of jars to you, whether you accepted Player 1's jars, and how many blue balls were drawn from your urns at the end of the round.



Wait 5 seconds before the next page can appear...

‹ 1 2 3 ›

Figure A6: Game Screen 3

(Page 4 of 10)

There are 2 types of urns: "High Blue" and "Low Blue". You will have one urn of each type in every round. The urns each have 100 balls at the beginning of each round, each ball is either blue or red. The chance of drawing blue balls from an urn is higher when there are more blue balls relative to red balls in that urn. You can change the chance of drawing blue balls from an urn at the end of the round by mixing the balls in one of the jars from Player 1 into that urn.

There are 2 types of jars: "Low Quality" and "High Quality". The quality of each jar is indicated by the number of blue balls: it can range from 0 to 100. When mixed into an urn, a jar can increase or decrease the average chance of drawing blue balls from the urn at the end of the round. With mixing, the number of blue balls in an urn at the end of the round will be equal to the sum of the number of blue balls in the jar and the number of blue balls in that urn (instead of the number of blue balls in that urn at the beginning of the round). For example, if the number of blue balls in a jar is higher than the number of blue balls in an urn, mixing the balls from that jar into that urn will increase the chance of drawing blue balls from that urn at the end of the round.

Wait 5 seconds before the next page can appear...

‹ 1 2 3 4 ›

Figure A7: Game Screen 4

(Page 5 of 10)

In each round, you can see the type and the actual number of blue balls in Player 1's jars as well as the type and actual number of blue balls in each of your urns before making any decisions in each round.

Unlike you, Player 1 will not be able to observe the exact number of blue balls in the jars that he/she received but he/she will be told whether he/she received a pair of identical "High Quality" or a "Low Quality" jars for that round. A "High Quality" jar has a number of blue balls (out of 100) that can be equal to any number between 70-100 with equal chance, whereas a "Low Quality" jar has a number of blue balls (out of 100) that can be equal to any number between 0-70 with equal chance. Player 1 also knows that you have a "High Blue" urn as well as a "Low Blue" urn. Player 1 also knows that a "High Blue" urn has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 40-100 with equal chance while a "Low Blue" urn has a number of blue balls (out of 100) (before mixing) that can be equal to any number between 0-60 with equal chance. Unlike you, Player 1 will neither observe nor receive any signals about the exact number of blue balls of either of your urns.

You, Player 2, can see the actual number of blue balls in the jars as well as the actual number of blue balls in each of your urns before making any decisions in each round.

Wait 5 seconds before the next page can appear...

‹ 1 2 3 4 5 ›

Figure A8: Game Screen 5

(Page 6 of 10)

Instructions for Player 2

In each of the 10 rounds, if Player 1 offered you the jars, you can decide whether to:

- Decline both of Player 1's jars

  OR

- Mix all the balls from a jar offered by Player 1 into in your High Blue urn and decline the other jar

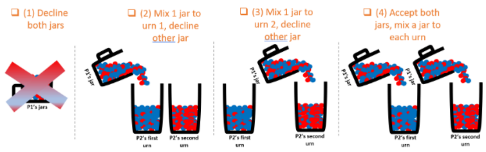
  OR

- Mix all the balls from a jar offered by Player 1 into in your Low Blue urn and decline the other jar

  OR

- Mix all the balls from a jar offered by Player 1 into in your High Blue urn and mix all the balls from the other jar offered by Player 1 into in your Low Blue urn

If Player 1 did not make an offer of jars in a round, you will not move for that round and will earn zero for that round. You, Player 2, cannot mix the balls from both jars into a single urn.



Wait 5 seconds before the next page can appear...

‹ 1 2 3 4 5 6 ›

Figure A9: Game Screen 6

After these introduction screens, four more screens will be shown to outline how subjects will be paid and to give subjects a chance to see how the stage-game interfaces for them and the player they are playing against look like.

First, subjects will see a verbal description of the incentive scheme associated with their assigned treatment condition. Figure A10 shows this screen for status quo and Figure A11 shows it for holistic.

(Page 7 of 10)

Your goal in the game is to minimize the percentage of red balls among balls drawn from urns where mixing occurred.

Player 1 earns by offering and getting more jars accepted, gets a penalty if both of his/her jars was declined, and gets zero if he/she made no offer or if only one of the jars were accepted.

Player 2 earns by mixing balls from jars to urns. He/she earns more by mixing more jars but he/she is penalized for for each red ball was drawn from every urn that was mixed with a jar. Player 2 gets zero if Player 1 made no offer.

We provide more details on the payment schemes for both Players in the next screen.

Wait 5 seconds before the next page can appear...

‹ 1 ... 4 5 6 7 8 ›

Figure A10: Game Screen 7 for Status Quo

(Page 7 of 10)

The aim of the game is to maximize number of blue balls drawn from all the urns.

Both Player 1 and Player 2 get rewards based on the number of blue balls and penalties based on the number of red balls drawn at the end of each round.

We provide more details on the payment schemes for both Players in the next screen.

Wait 5 seconds before the next page can appear...

‹ 1 2 3 4 5 6 7 ›

Figure A11: Game Screen 7 for Holistic

Next, subjects will see a schematic and a more detailed verbal description of how payoffs are determined. It clearly depicts how payoff levels for the subject and their counterpart in the game are determined based on each other's actions. Figure A12 shows this screen for status quo and Figure A13 shows it for holistic.

(Page 8 of 10)

Player 1 earns tokens by offering two jars that were subsequently BOTH accepted by Player 2, gets a penalty by offering two jars that was subsequently BOTH declined by Player 2, gets 0 tokens by offering two jars and only ONE jar was accepted by Player 2, and gets 0 tokens by not making an offer at all.

- If Player 1 did NOT OFFER the pair of jars to Player 2 in a round of the study, both players will receive 0 tokens for that round.
- If Player 1 offered the pair of jars to Player 2 in a round of the study and Player 2 DECLINED BOTH jars, 30 tokens are deducted from Player 1's Bank and Player 2 will receive 0 tokens for that round.
- If Player 1 offered the pair of jars to Player 2 in a round of the study and Player 2 ACCEPTED ONLY ONE jar (that he/she has to mix into either one of the urns), Player 1 gets 0 tokens for that round
- If Player 1 offered the pair of jars to Player 2 in a round of the study and Player 2 ACCEPTED BOTH jars (that he/she has to mix into the two urns), Player 1 gets 10 tokens for that round

Player 1 has clear incentives to offer a pair of jars if Player 2 will accept both jars and to not offer a jar if Player 2 will decline both jars

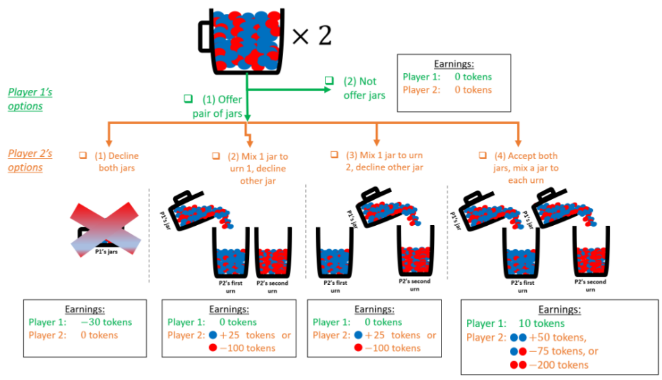
Player 2 can only earn tokens by mixing balls from jars into urns.

- If Player 2 accepted a jar, he/she has to mix it with one of the urns.
- If Player 2 accepted both jars, he/she has to mix them both, one jar with each of the urns.
- If Player 2 did not perform a mixing, either because no jar was offered to him/her or if he/she chose not to decline both jars, Player 2 gets 0 tokens.

- Player 2 earns 25 tokens by mixing a jar with an urn and drawing a blue ball from the mixed urn.
- Player 2 loses the earnings and instead receive an 100 token penalty that is deducted from his/her Bank if a red ball was drawn from the mixed urn.
- Player 2 will neither get any additional rewards or penalties for the urn he/she did not mix.

  In other words, Player 2 gets:

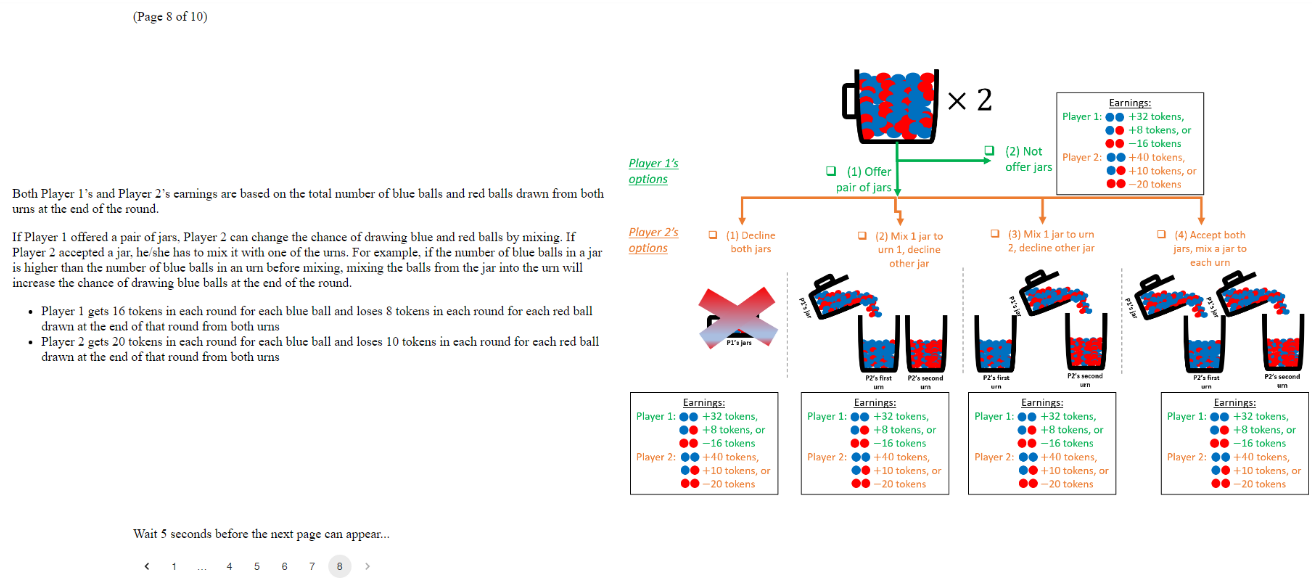
  - +50 tokens if he mixed both jars and blue balls were drawn from BOTH urns
  - -75 tokens if he mixed both jars and a blue balls were drawn from one urn and a red ball from the other urn
  - -200 tokens if he mixed both jars and red balls were drawn from BOTH urns
  - +25 tokens if he mixed one jar with one of the urns and a blue ball was drawn from that urn
  - -100 tokens if he mixed one jar with one of the urns and a red ball was drawn from that urn
  - 0 tokens if he declined the jars and did no mixing



Wait 5 seconds before the next page can appear...

‹ 1 ... 5 6 7 8 9 ›

Figure A12: Game Screen 8 for Status Quo


(Page 8 of 10)

Both Player 1's and Player 2's earnings are based on the total number of blue balls and red balls drawn from both urns at the end of the round.

If Player 1 offered a pair of jars, Player 2 can change the chance of drawing blue and red balls by mixing. If Player 2 accepted a jar, he/she has to mix it with one of the urns. For example, if the number of blue balls in a jar is higher than the number of blue balls in an urn before mixing, mixing the balls from the jar into the urn will increase the chance of drawing blue balls at the end of the round.

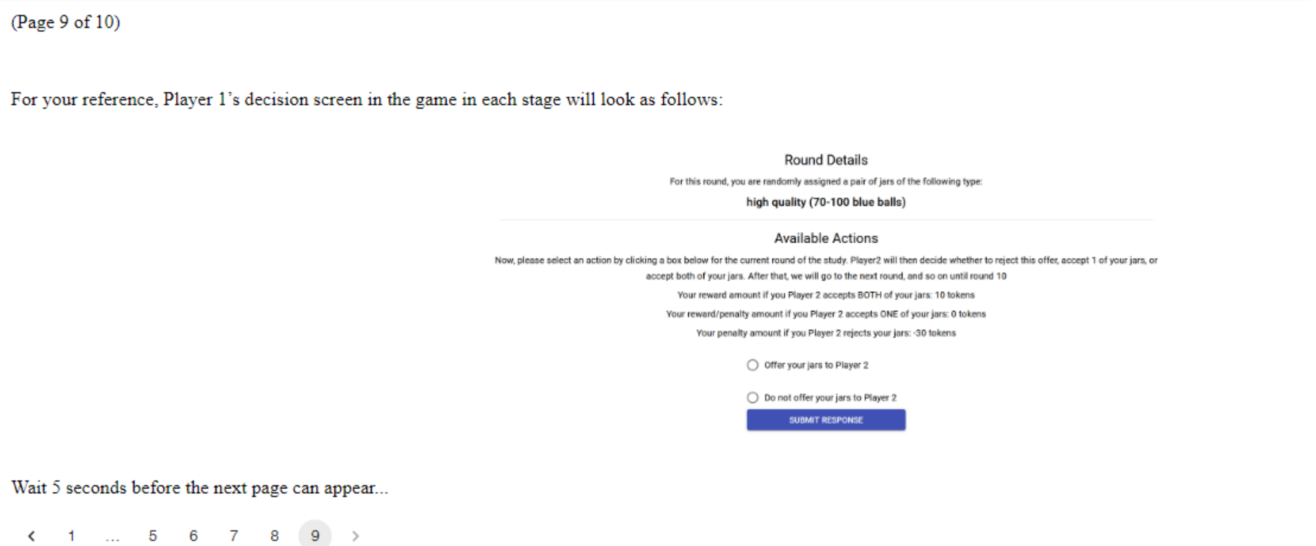
- Player 1 gets 16 tokens in each round for each blue ball and loses 8 tokens in each round for each red ball drawn at the end of that round from both urns
- Player 2 gets 20 tokens in each round for each blue ball and loses 10 tokens in each round for each red ball drawn at the end of that round from both urns

Wait 5 seconds before the next page can appear...

‹ 1 ... 4 5 6 7 8 ›

Figure A13: Game Screen 8 for Holistic

Then, subjects will see what they would expect to see as the game interface in the stage game, as well as what the other play would see. Player 1's stage game screen is always shown first and then Player 2's. Figures A14 and A15 shows these screens for status quo and Figures A16 and A17 shows them for holistic.


(Page 9 of 10)

For your reference, Player 1's decision screen in the game in each stage will look as follows:

Round Details

For this round, you are randomly assigned a pair of jars of the following type:

high quality (70-100 blue balls)

Available Actions

Now, please select an action by clicking a box below for the current round of the study. Player2 will then decide whether to reject this offer, accept 1 of your jars, or accept both of your jars. After that, we will go to the next round, and so on until round 10

Your reward amount if you Player 2 accepts BOTH of your jars: 10 tokens

Your reward/penalty amount if you Player 2 accepts ONE of your jars: 0 tokens

Your penalty amount if you Player 2 rejects your jars: -30 tokens

- Offer your jars to Player 2
- Do not offer your jars to Player 2

SUBMIT RESPONSE

Wait 5 seconds before the next page can appear...

‹ 1 ... 5 6 7 8 9 ›

Figure A14: Game Screen 9 for Status Quo

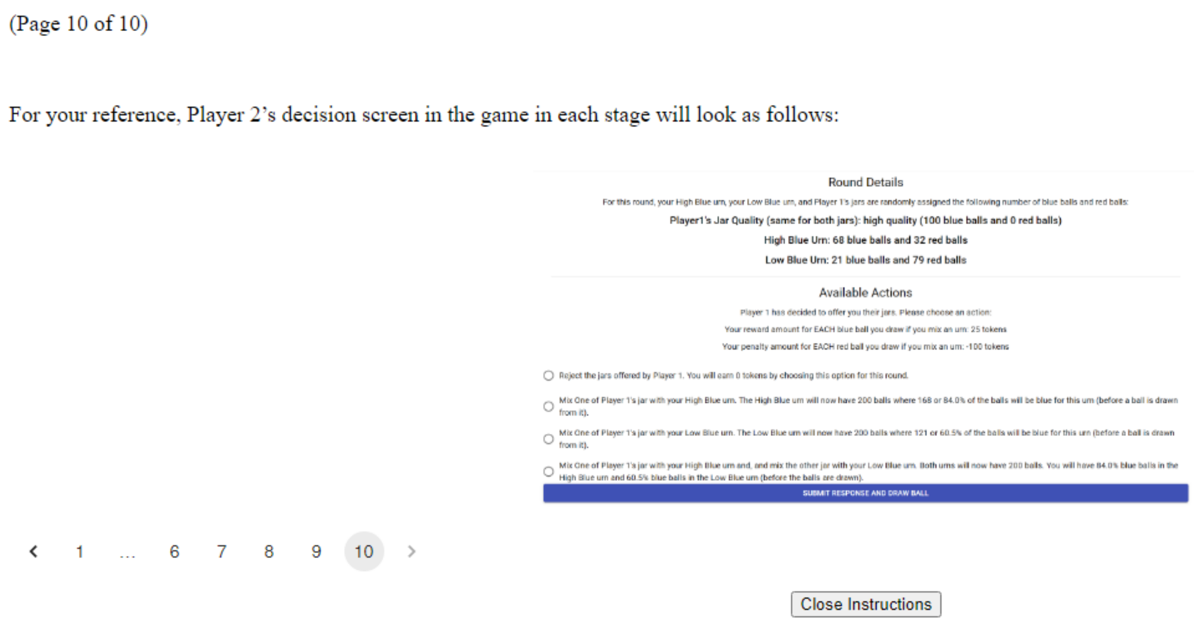


Figure A15: Game Screen 10 for Status Quo

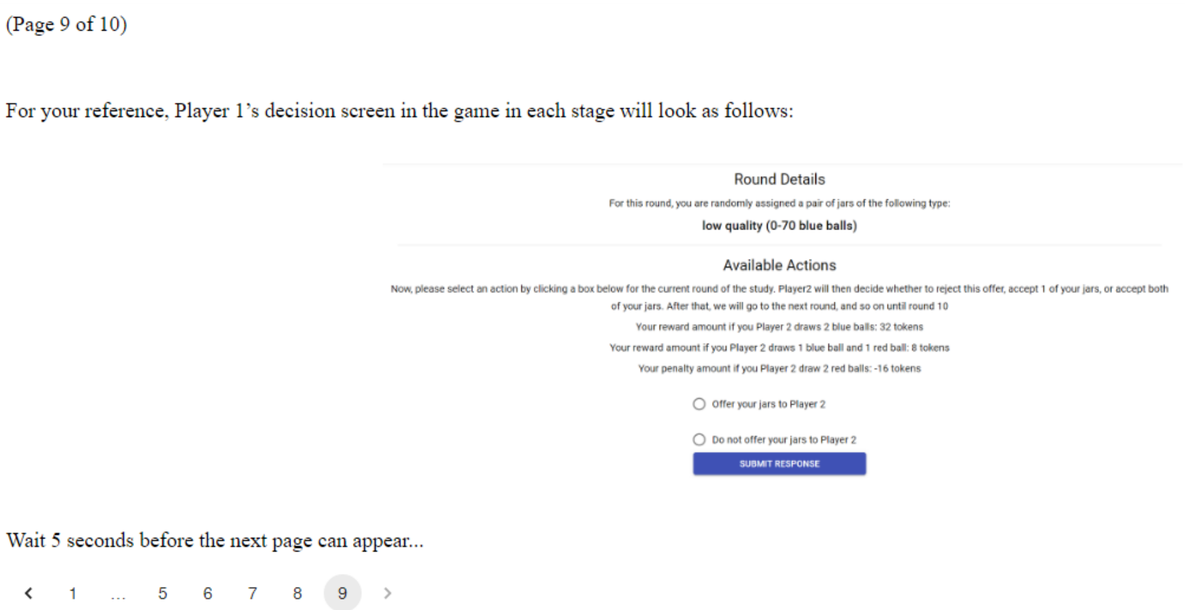


Figure A16: Game Screen 9 for Holistic

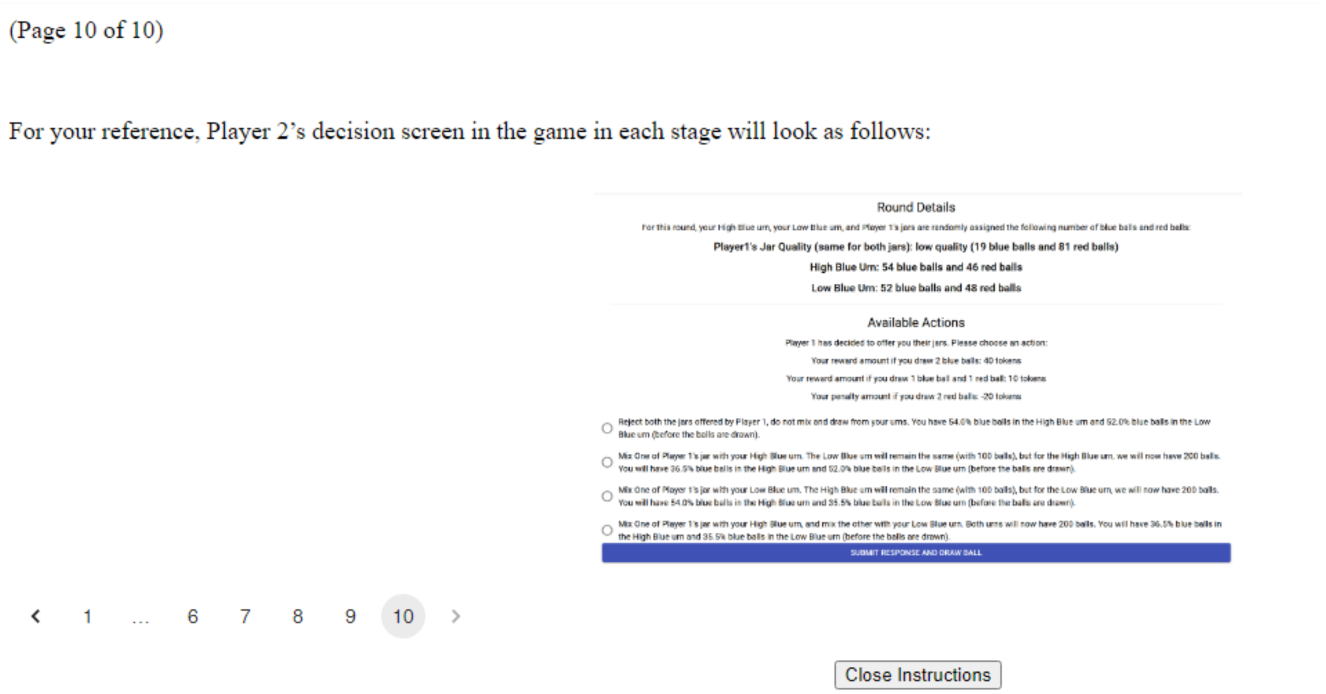


Figure A17: Game Screen 10 for Holistic

A.1.4 Comprehension Quiz

To reinforce comprehension of the rules of the game, the introduction screens are followed by five comprehension questions before the actual game play commences. The correct answers, along with a pop-up text box to explain why such answers are correct, to these questions are provided to the subjects if they got a question wrong. The subject will not be able to advance until they correctly answered all the five questions. The order of these questions are randomized. The questions are shown in the figures below.

If the question shown in Figure A18 is answered incorrectly, the following answer is shown for both conditions:

- "Answer is 40%. Since the jar and the urn has the same number of balls in total and the jar has 50% blue balls and the urn has 30% blue balls, the percentage of blue balls will be the average of the 2, or 40%. Another way to see this is that 50 blue balls from jar + 30 blue balls from urn = 80 blue balls in total. Dividing this by 200 balls (100 balls from jar and 100 balls from urn), we get 40%."

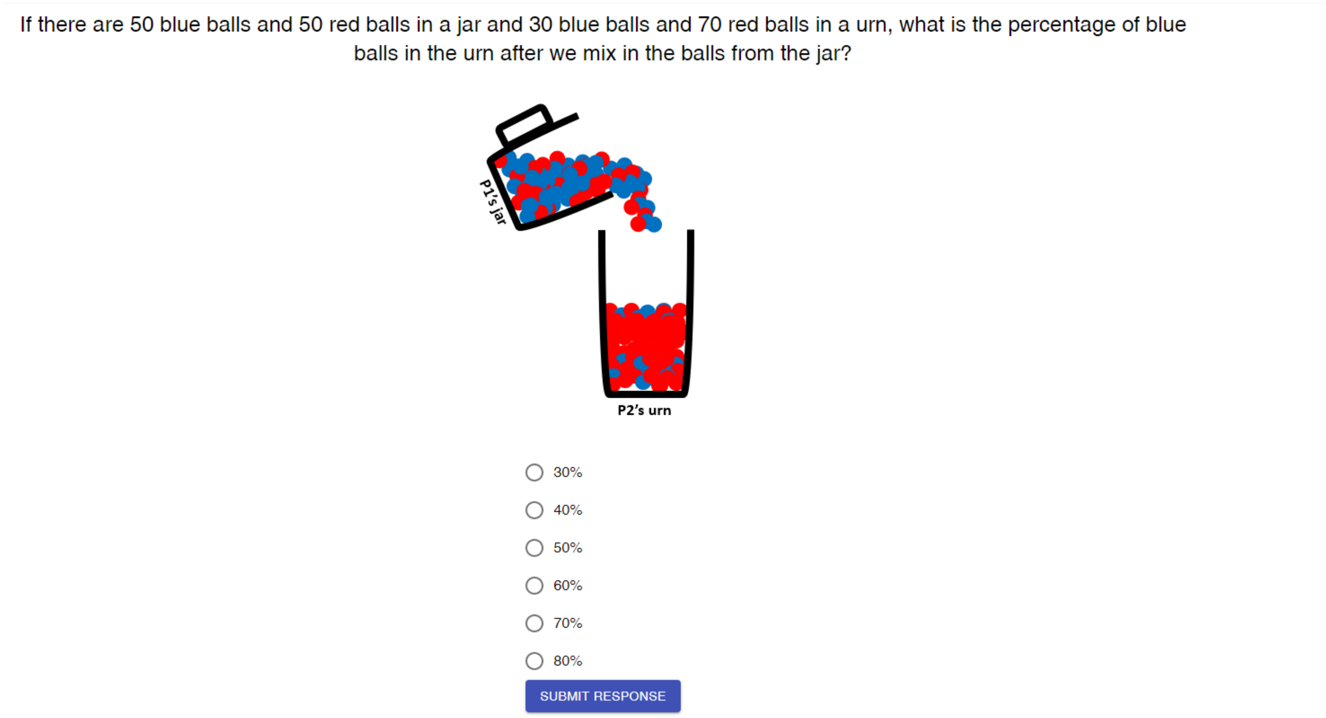


Figure A18: Comprehension Question 1

If the question shown in Figure A19 is answered incorrectly, the following answer is shown for both conditions:

- "Answer is Higher than before mixing. Before mixing, the urn has a lower percentage of blue balls. Since the jar has a higher percentage of blue balls, mixing will increase the percentage of blue balls in the urn. So if the one wants to have the highest chance of drawing a blue ball, one would mix."

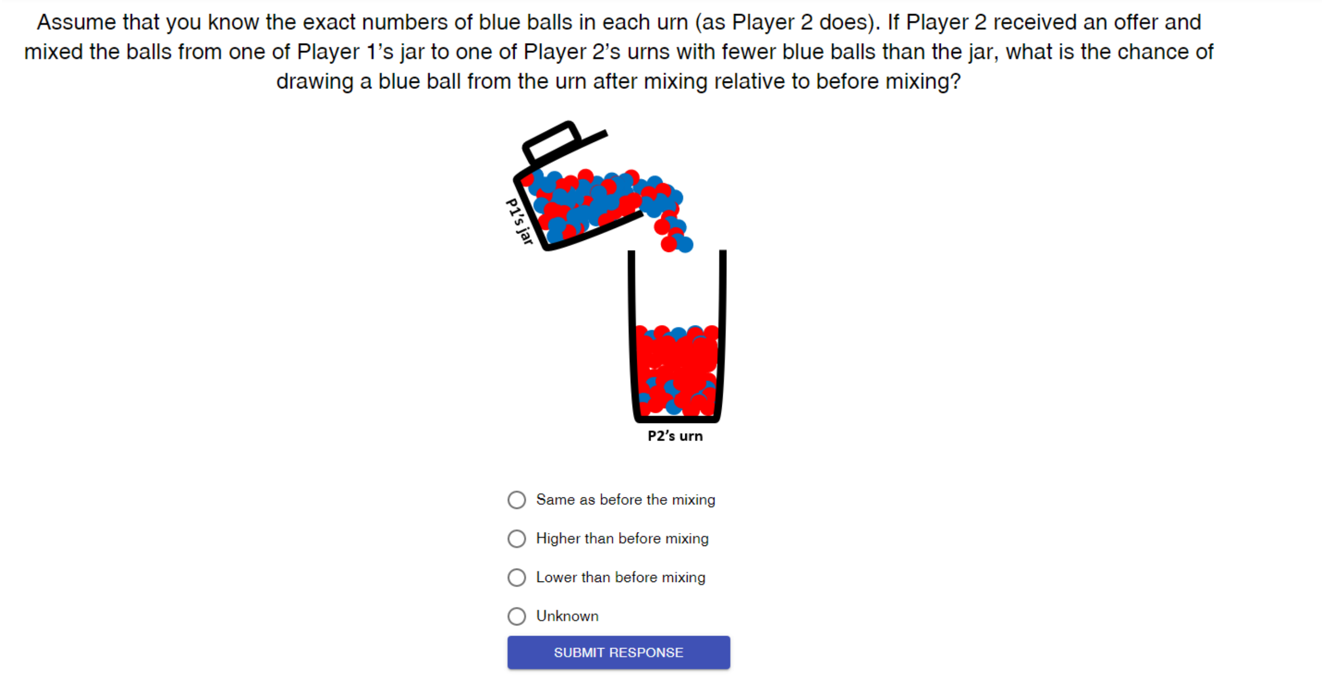


Figure A19: Comprehension Question 2

If the question shown in Figure A20 is answered incorrectly, the following answer is shown depending on the treatment condition:

- For Status Quo: "Answer is Unclear with given information. It depends whether the mixed urn(s) will provide good enough odds for Player 2. For example, if Player 2 is given jars with 99 blue balls and 1 red ball, the mixed High Blue urn will still have a 99.5% chance of drawing a blue ball. While 99.5% is lower than 100% , Player 2 might still find it worth his/her while to make the bet. On the other hand, if the jar has fewer blue balls, say 1 blue ball and 99 red balls, mixing it with the urn with 100 blue balls will now lower the chance of drawing a blue ball from 100% to 50.5% , drastically lowering the chance of drawing a blue ball - while some more risk loving Player 2 might still want to mix for a chance to win some earnings, some others might rather not take the gamble. Also, note that Player 2 will never mix the jar with the urn with 100 red balls as mixing any given jar with the urn with 100 blue balls will give him/her better odds of drawing blue balls."

- For Holistic: "Answer is Reject any offer from Player 1 and Mix the balls from Player 1's jar with the urn with 100 blue balls. Player 2 will never reject any jar offer that can at least improve the odds of one of his/her urns. Given that the urn with 100 red balls has 0% chance of drawing a blue ball, any chance to add a blue ball in the mix will be an improvement - therefore he/she will never reject a jar. Similarly, with whatever jar (not 100 red balls or 100 blue balls) that Player 2 received, he/she can only make the odds of drawing a blue ball from the urn with 100 blue balls worse, as the current chance is 100%. Therefore, he/she will never mix a jar with the urn with 100 blue balls instead of the urn with 100 red balls."

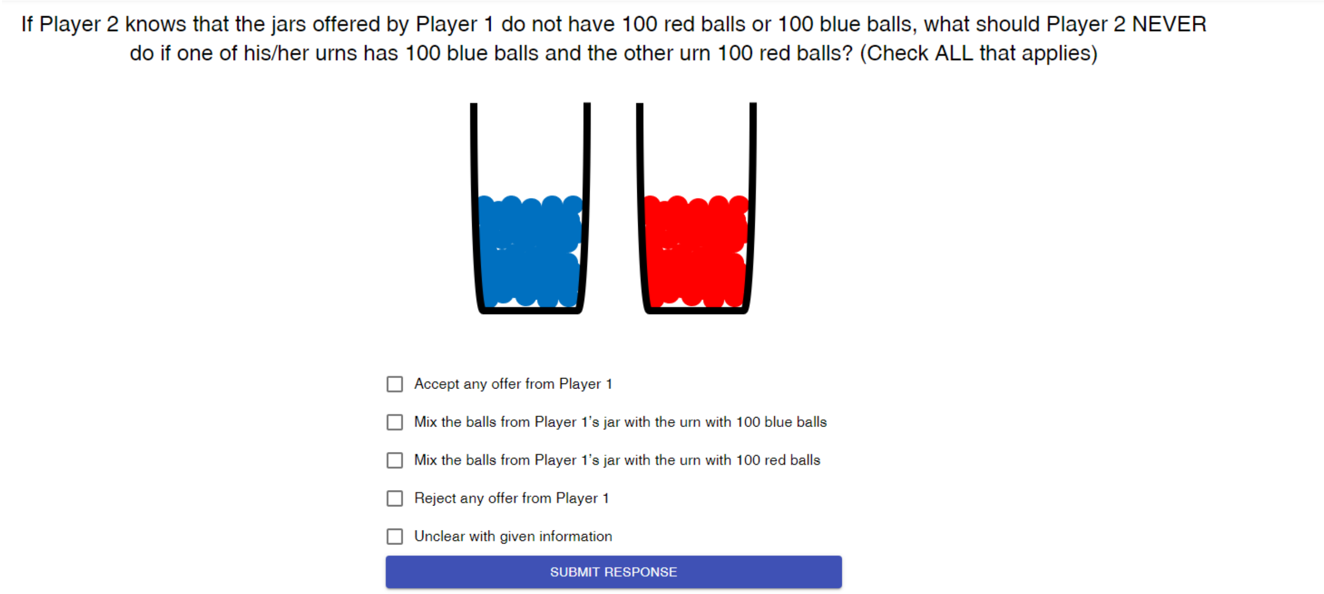


Figure A20: Comprehension Question 3

If the question shown in Figure A21 is answered incorrectly, the following answer is shown for both conditions:

- "Answer is No. The jar has 49% blue balls and the urn has 50% blue balls. Mixing a jar with lower percentage of blue balls that the urn will only lower the percentage of blue balls in the urn. So if the one wants to have the highest chance of drawing a blue ball, one would not mix and just draw from the urn."

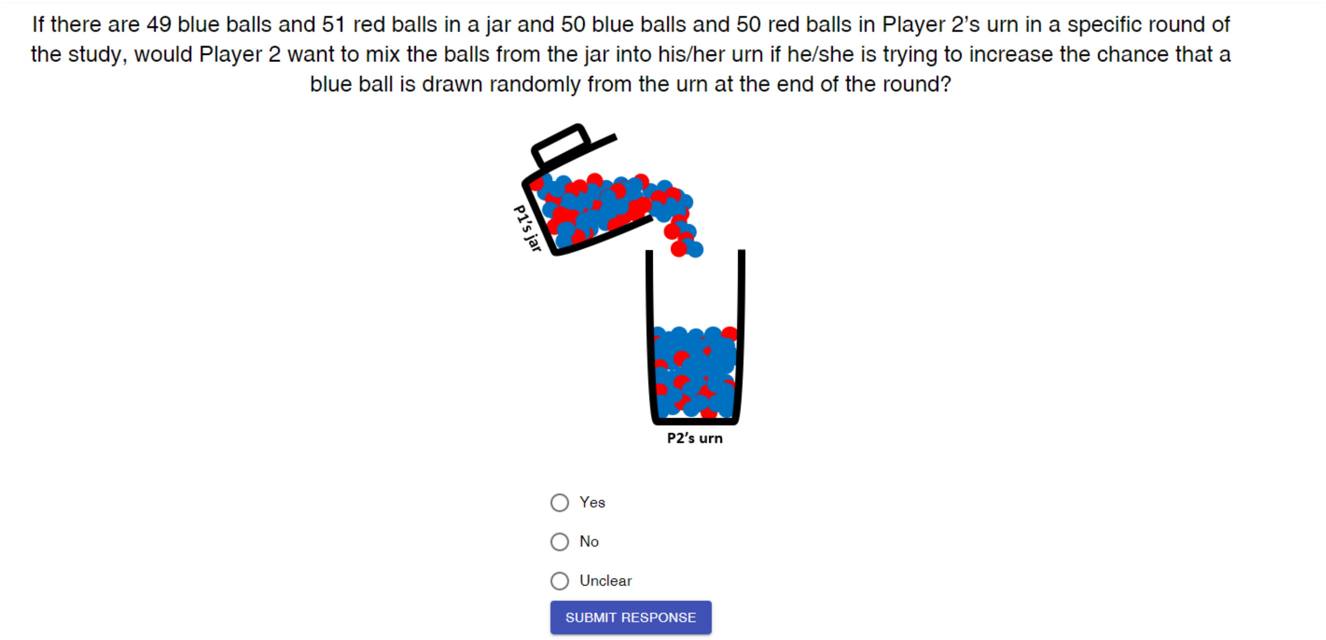


Figure A21: Comprehension Question 4

If the question shown in Figure A20 is answered incorrectly, the following answer is shown depending on the treatment condition:

- For Status Quo: "Answer is Definitely not make an offer to Player 2. This is because Player 1 will lose tokens if he/she made an offer to Player 2 and it was not accepted. Therefore, if Player 1 knows that the jars will be rejected, he/she will never offer it in the first place."

- For Holistic: "Answer is Unclear/Do not know. Player 1's payoff depends on what color balls are drawn from the 2 urns. While he/she can influence the odds of drawing blue balls by offering Player 2 a pair of jars as an option of changing the percentage of blue balls for one or both of his/her urns, if Player 2 were to not accept the jars for sure Player 1 loses nothing. However, if Player 1 is not completely sure if Player 2 would accept the jars. If Player 1 is not 100% sure that Player 2 would accept both jars but 99% sure that Player 2 would reject the jars, it is still in Player 1's interest to offer the jars in the off chance (1% chance) that Player 2 might be able to accept and improve the odds of drawing a blue ball from one or both of the urns."

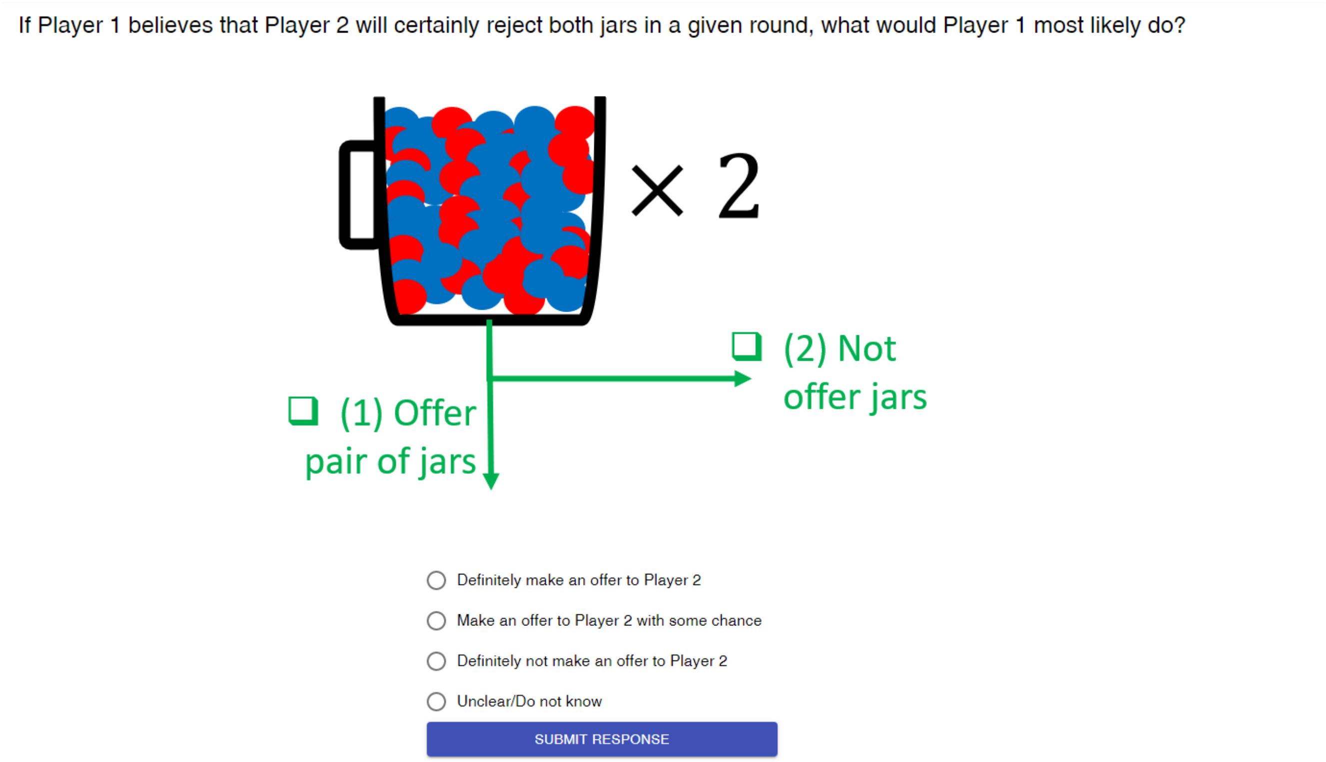


Figure A22: Comprehension Question 5

A.1.5 Stage Game Screens

Finally, the subjects play the stage game for 10 rounds.

Player 1 gets action options (in a multiple choice format) and a reminder that shows them the level of payoffs he should expect to receive based on Player 2's and their own actions (and the draw of the balls if applicable). This decision screen for Player 1s in the status quo condition is shown in A23 and shown in A24 for holistic.

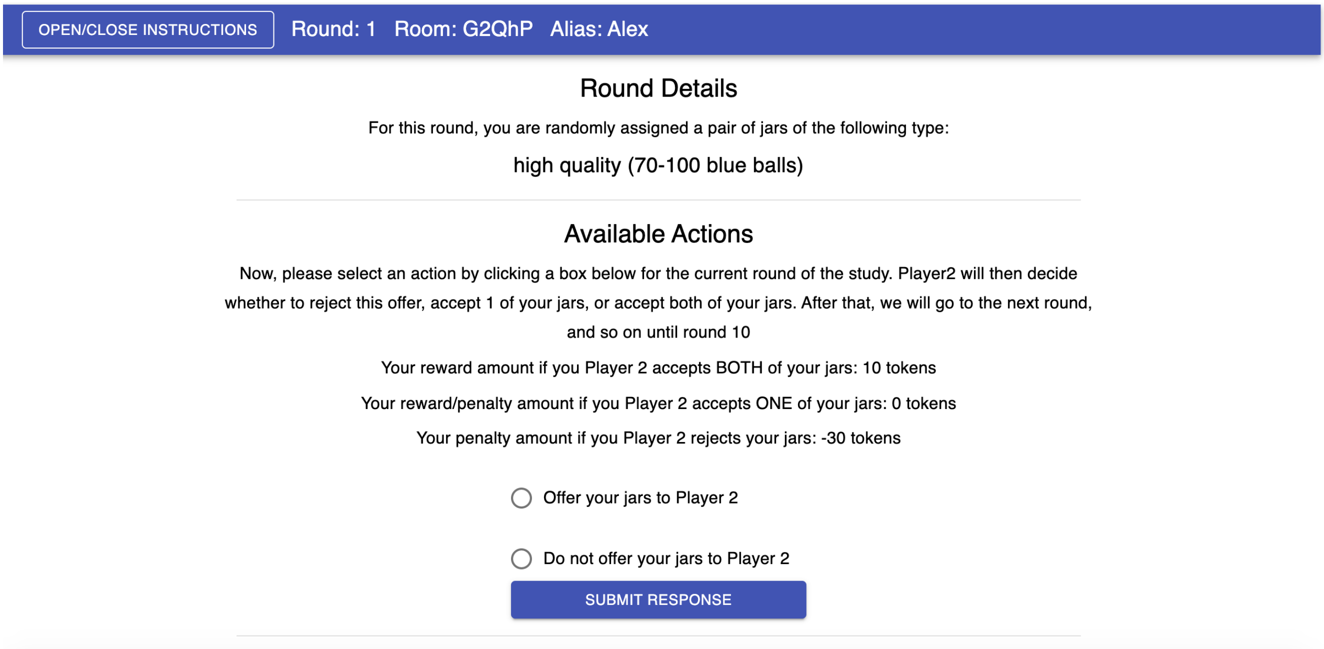


Figure A23: Player 1 Stage Game Decision Screen for Status Quo

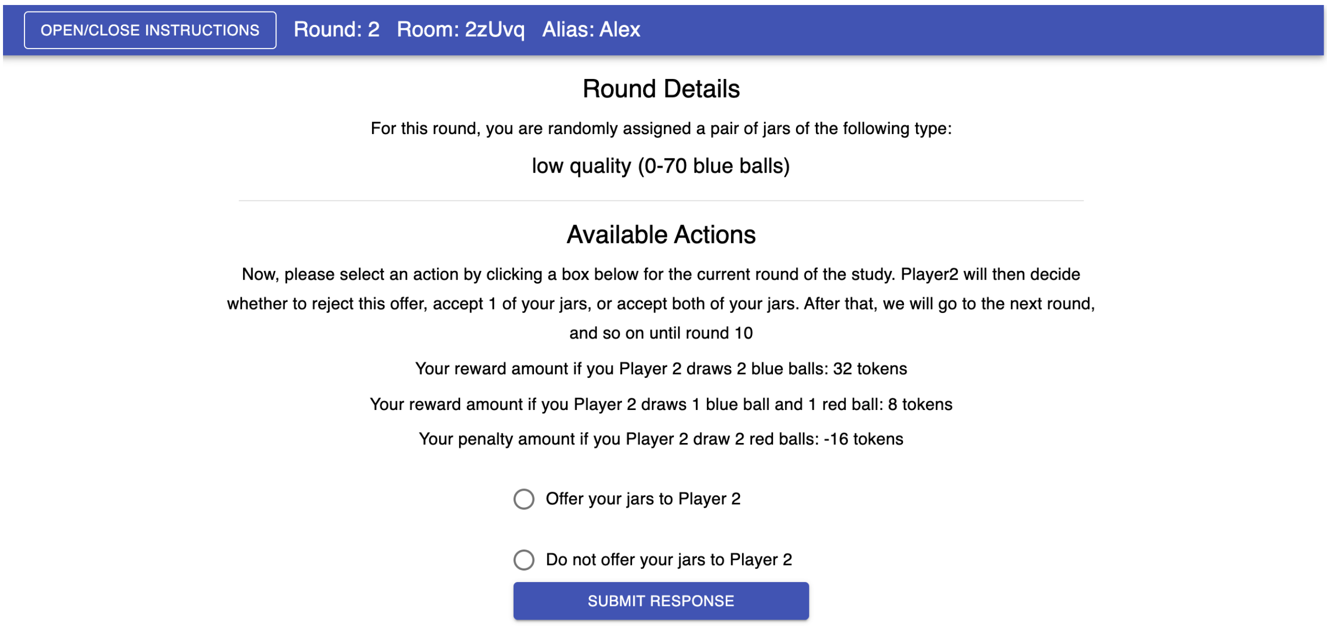


Figure A24: Player 1 Stage Game Decision Screen for Holistic

After Player 1 moves, Player 2 will get to move in the round. If Player 1 offered Player 2 the jars, Player 2 sees a screen that show her action options and a reminder that shows her the level of payoffs she should expect to receive based on the draw of a ball from the urns at the end of the round. Note that along with each action option for Player 2, the expected number of blue/red balls associated with the urn post-mixing is also provided to ease the need for Player 2 to do the calculations. This decision screen for Player 2s in the Status Quo condition is shown in A25 and shown in A26 for those in the Holistic condition.

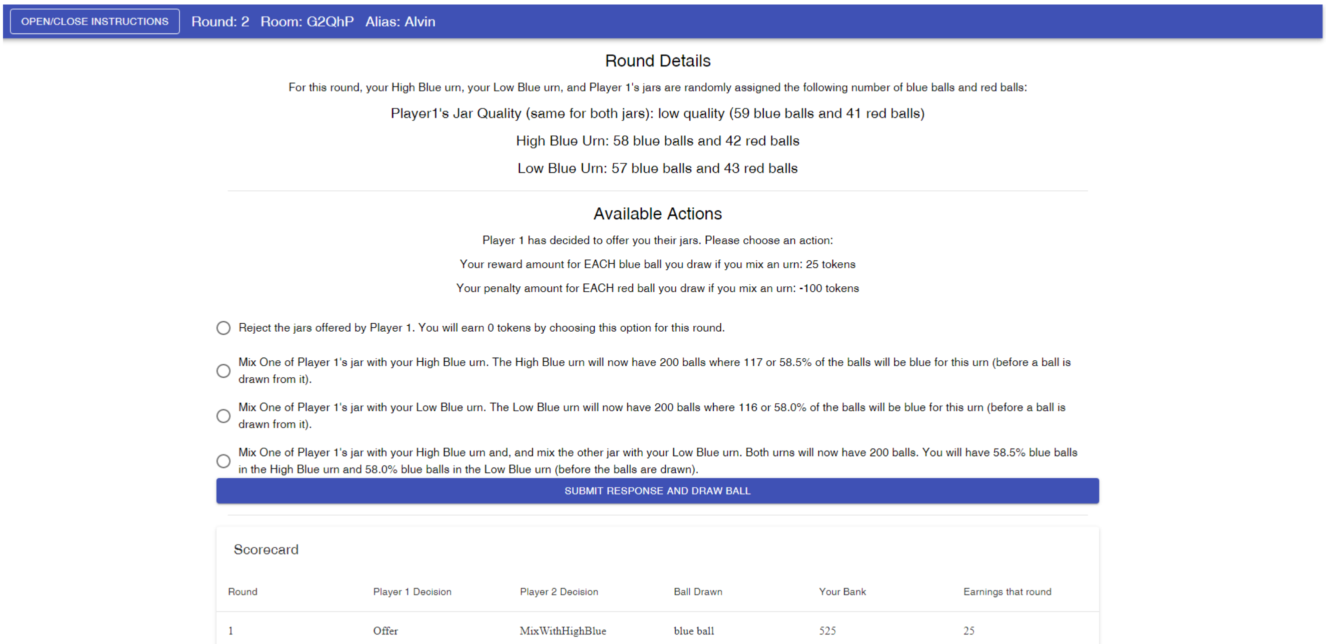


Figure A25: Player 2 Stage Game Decision Screen for Status Quo

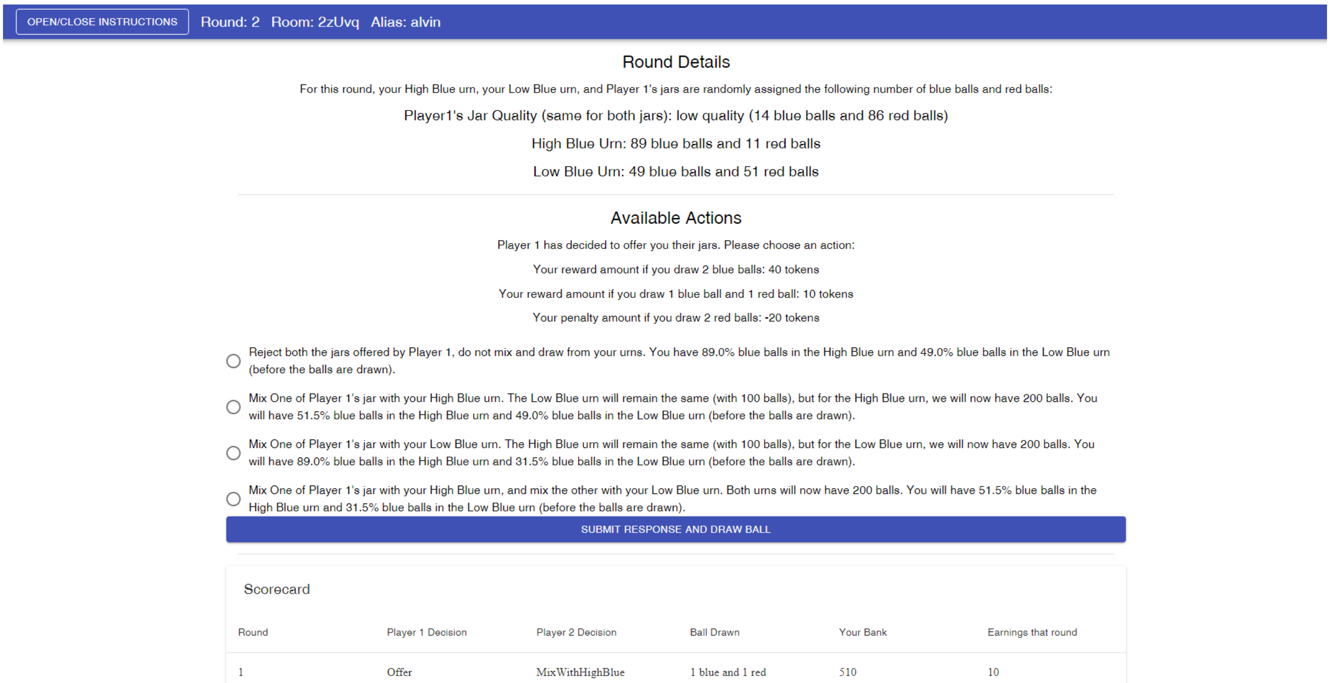


Figure A26: Player 2 Stage Game Decision Screen for Holistic

After both players have made their decisions for each round, a screen showing the actual ball(s) drawn from the urn(s) at the end of the round and the earnings for the player in the round is shown. A table summarizing the actions taken by each player, actual ball drawn, and each player's earnings and bank in each previous round is shown at the bottom of this page. This table is also shown in the bottom of each decision screen. The post-decision screen summarizing actions, balls drawn and payoffs for each round is shown below (Figure A27 for Player 1 and Figure A28 for Player 2.).

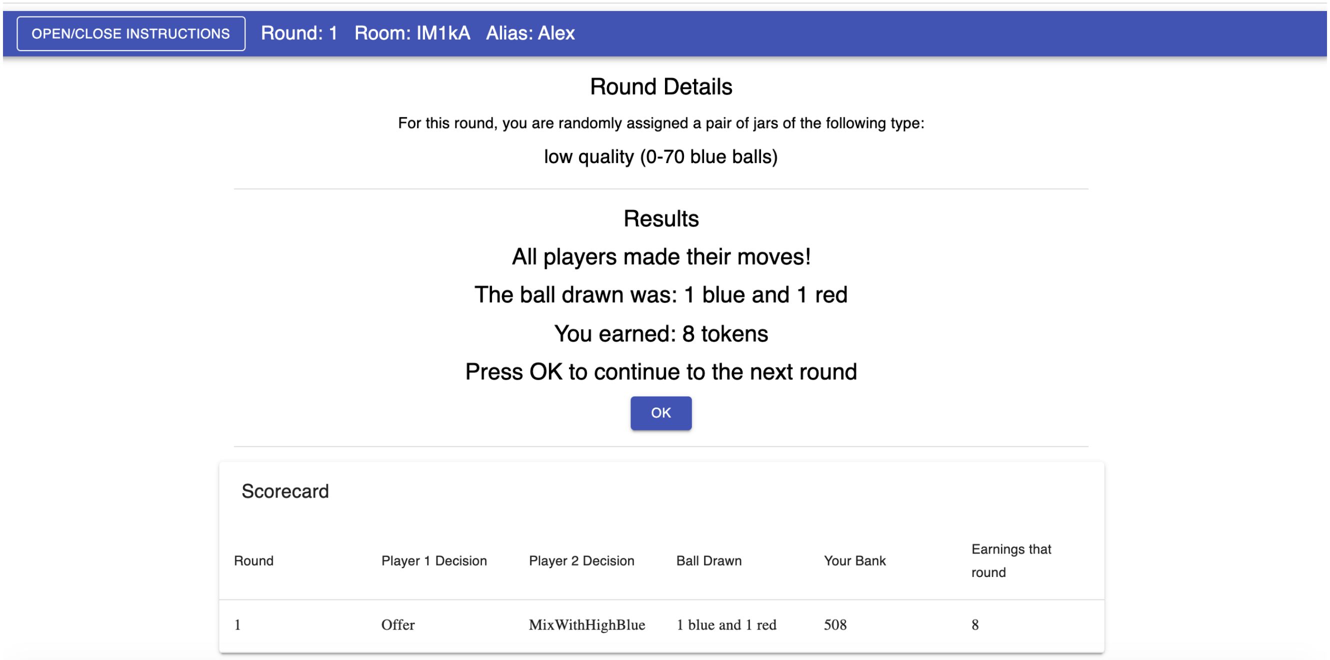


Figure A27: Player 1 Stage Game Post-Decision Screen

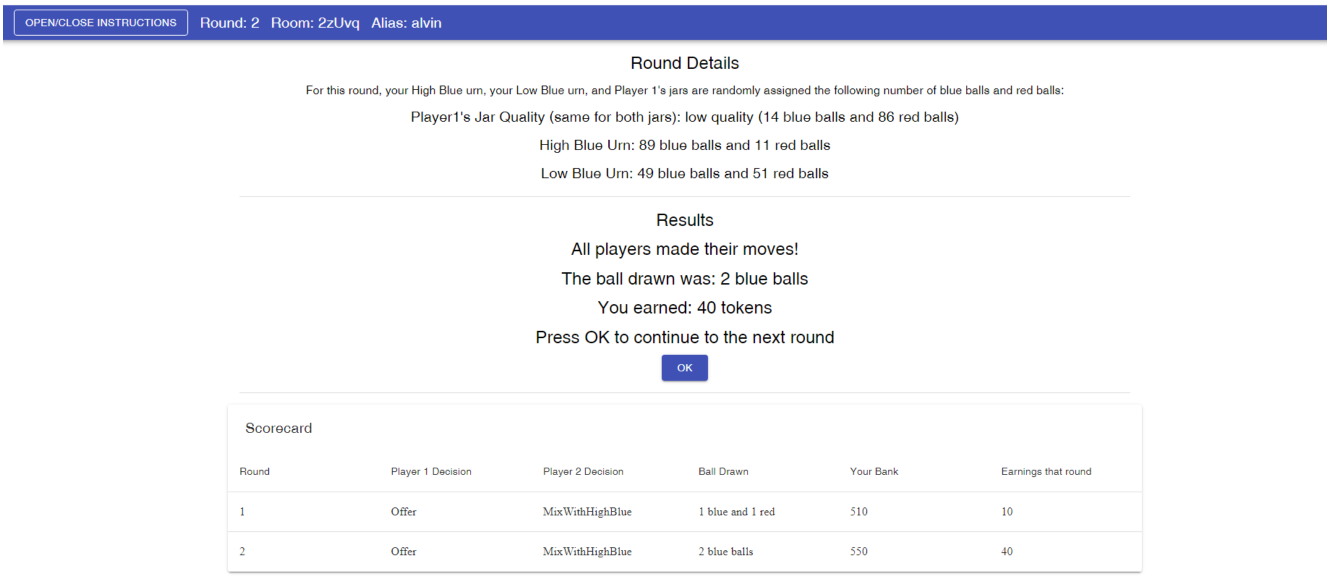


Figure A28: Player 2 Stage Game Post-Decision Screen

A.1.6 Demographic Information Collection

We constructed the survey tool to gather demographic information using Qualtrics software for customers who clicked the URL on Prolific to participate in the experiment. Upon opening the link, respondents must read and consent to continue (see Figure A29).

Protocol Director: Professor Alvin Roth

Protocol Number: IRB-44868 IRB
Approval Date: March 31, 2021
Expiration Date: March 31, 2022

Description: You are invited to participate in a research study on decision-making. You will be asked to read several pages of instructions. Then you will be asked to make several choices that will determine the precise amount you will be paid, and then possibly answer several survey questions.

Risk and benefits: We cannot and do not guarantee or promise that you will receive any benefits from this study. There are no risks associated with this study.

Time involvement: Your participation in this experiment will take approximately as long as is indicated in the advertisement.

Payments: You will be compensated at the advertised rate.

Subject's rights: If you have read this form and have decided to participate in this project, please understand your participation is voluntary and you have the right to discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled. You have the right to refuse to answer particular questions. Your individual privacy will be maintained in all published and written data resulting from the study.

Contact information:
Questions, concerns, or complaints: If you have any questions, concerns or complaints about this research study, its procedures, risk and benefits, you should ask the Protocol Director, Alvin Roth, (650)-725-9147.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at (650)-723-2480 or toll free at 1-866-680-2906 or via email at irbnonmed@stanford.edu. You can also write to the Stanford IRB, Stanford University, Stanford, CA 94305-5401.

By continuing with this study, you are consenting to participate.
Please make a copy of this consent form for your own records. You can do so by right-clicking and selecting "print" in most browsers. If you cannot do so on your browser, please contact the protocol director for a copy of the consent form.

I consent, begin the study

I do not consent, I do not wish to participate

Figure A29: Consent screen

Then, the subjects are asked to answer a series of questions listed below before they are redirected to the game itself. These question gather information about the subject on their (in the order of appearance): race (Figure A30) and ethnicity (A31), sex (Figure A32), state of residence (Figure A33), age (Figure A34), employment status (Figure A35), and education (Figure A36).

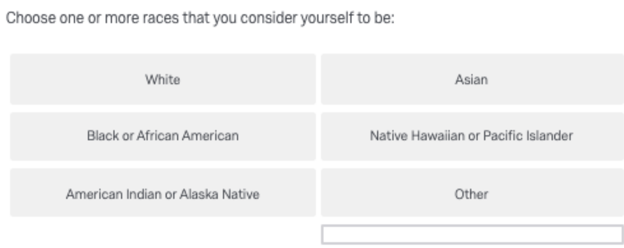


Figure A30: Survey Question about Subject's Race

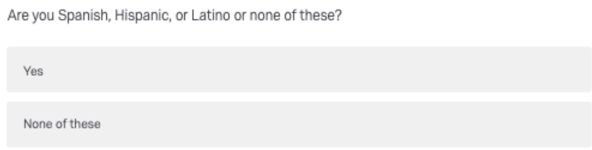


Figure A31: Survey Question about Subject's Ethnicity

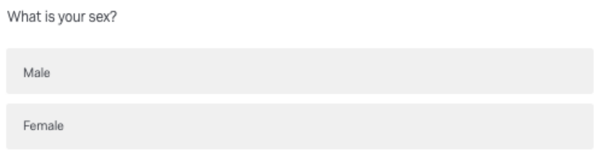


Figure A32: Survey Question about Subject's Sex

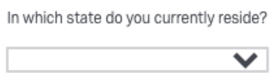
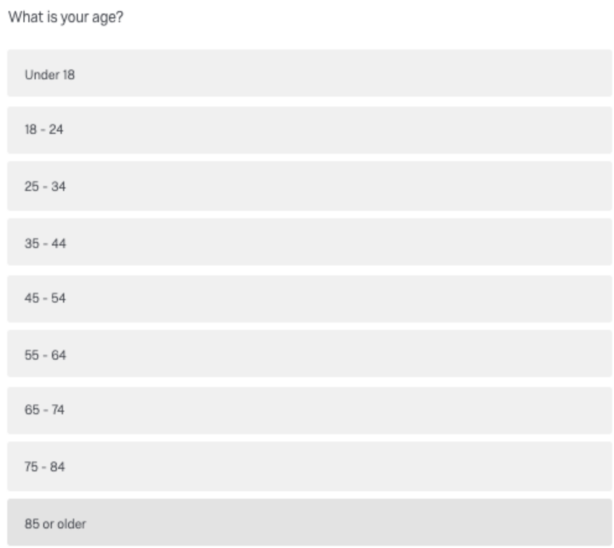


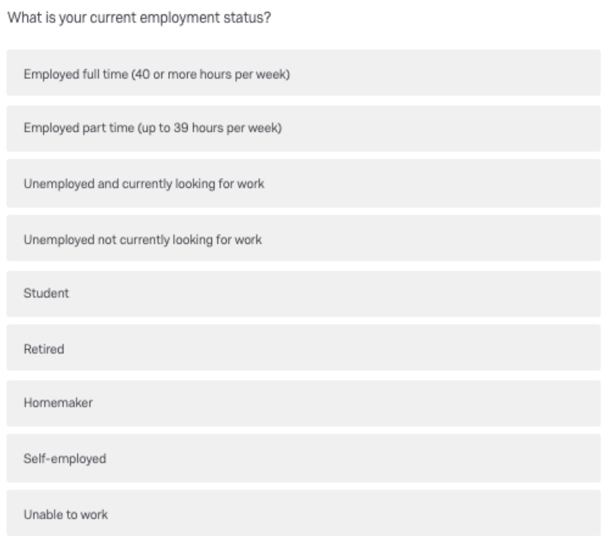
Figure A33: Survey Question about Subject's State of Residence



What is your age?

- Under 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- 55 - 64
- 65 - 74
- 75 - 84
- 85 or older

Figure A34: Survey Question about Subject's Age



What is your current employment status?

- Employed full time (40 or more hours per week)
- Employed part time (up to 39 hours per week)
- Unemployed and currently looking for work
- Unemployed not currently looking for work
- Student
- Retired
- Homemaker
- Self-employed
- Unable to work

Figure A35: Survey Question about Subject's Employment Status

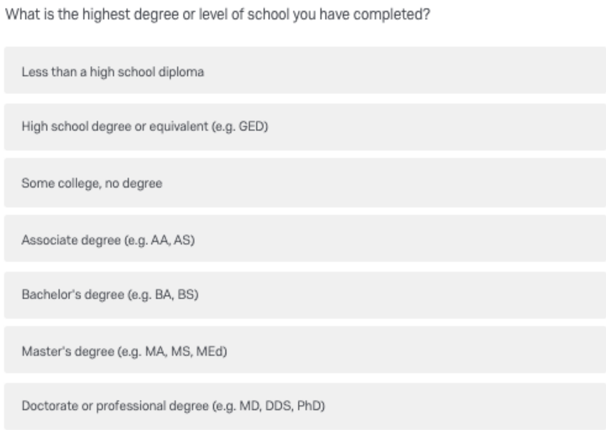


Figure A36: Survey Question about Subject's Education Status

After providing demographic information, the subjects will asked to submit their Prolific ID and will be automatically redirected to the game itself (see Figure A37).

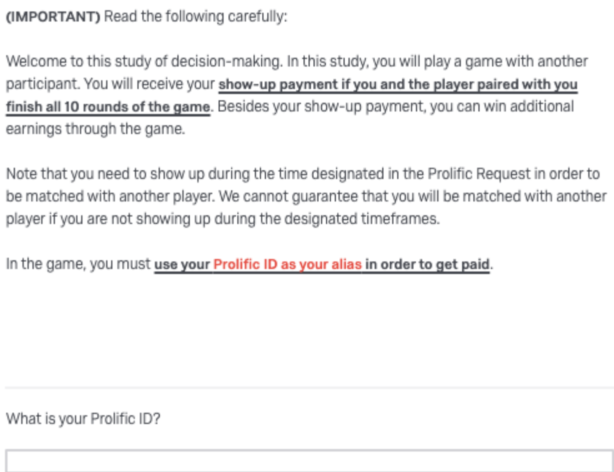
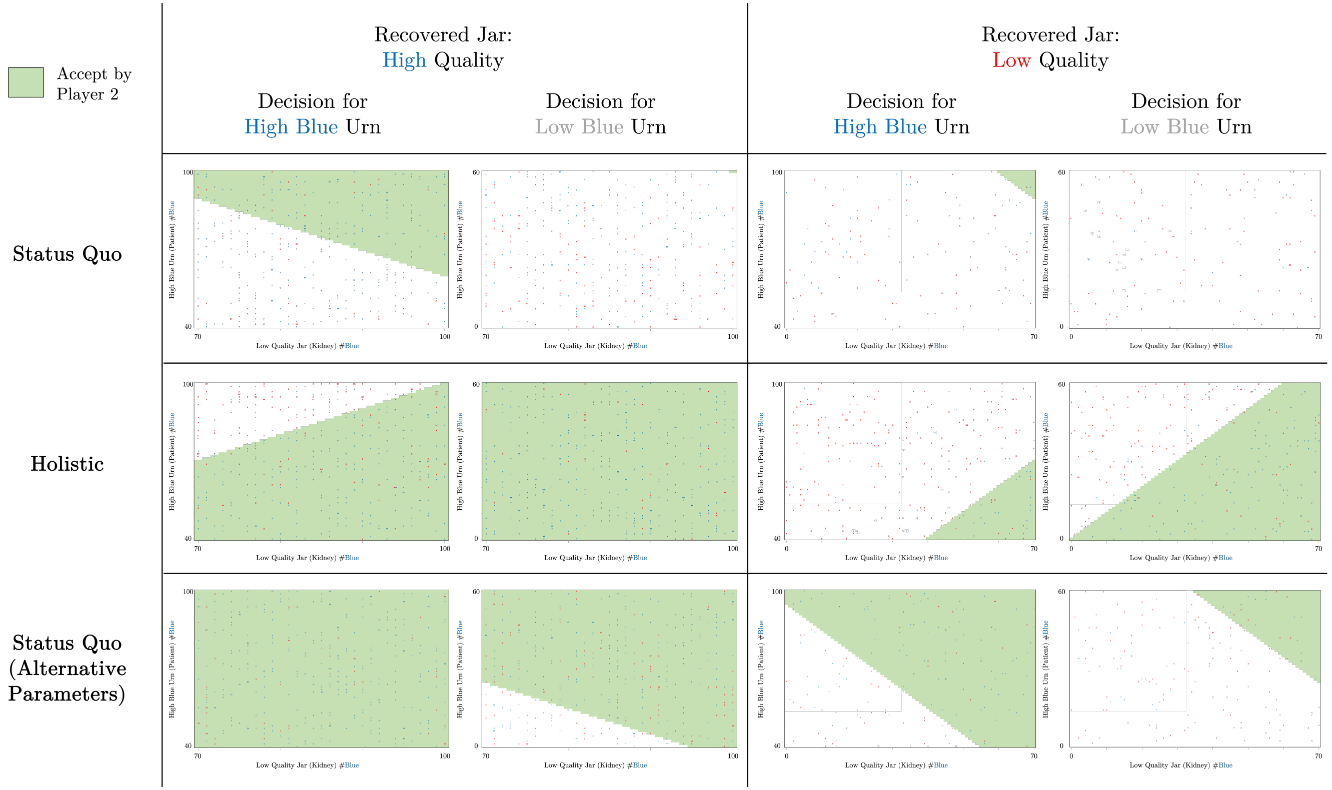


Figure A37: Transition Screen to Game

A.2 Supplementary Figure

Figure A38: Jar Acceptance Strategy for Player 2 (Transplant Center) for Various Manifested Jars and Urns With Actual Data



Notes: This figure illustrates Player 2's (TC's) optimal strategy as an expected income maximizer BUT UNLIKE Figure II, it has the actual experimental data plotted. An expected income maximizing Player 2 will only mix a jar into the urn if the proportion of blue balls in the mixed urn is high enough. Each grid in each (of the 12) graph above represent the number of blue balls (out of 200) that would arise with each possible combination of urn and jar given the respective types. The green shaded area represent the jar-urn combination that would yield an expected income Player 2 non-negative expected income: the top panel illustrate optimal Player 2 strategy under status quo treatment under the baseline experimental parameters, the middle panel illustrate this for both the holistic treatment under the baseline experimental parameters as well as Holistic Treatment 2 under the alternative parameters (optimal strategy same), and the bottom panel illustrate Player 2's optimal strategy under the Counterfactual Status Quo treatment under the alternative parameters. A point is plotted for each realized and offered jar across the experiment session: blue "Y" for an accepted jar, red "X" for a declined/discarded the jar. The actual data plotted on the top and middle panels are from the (respective treatment group data from) baseline experiment while the bottom panel shows data from the Counterfactual Status Quo treatment group under the alternative parameters.

A.3 Supplementary Tables

Table A1: Impact on Bad Outcomes (Red Balls Drawn)

	(1)	(2)	(3)	(4)	(5)	(6)
	% Red Balls for all Urns		% Red Balls for Low Blue Urns		% Red Balls for High Blue Urns	
Status Quo	0.034^{+}	0.031^{+}	0.026	0.031	0.041^{+}	0.032
	(0.018)	(0.0167)	(0.024)	(0.023)	(0.022)	(0.021)
Jar # Blue Balls		-0.002^{**}		-0.003^{**}		-0.001^{+}
		(0.000)		(0.000)		(0.000)
High Urn # Blue Balls		-0.004^{**}		-0.000		-0.007^{**}
		(0.000)		(0.001)		(0.001)
Low Urn # Blue Balls		-0.004^{**}		-0.008^{**}		0.000
		(0.000)		(0.001)		(0.001)
Constant	0.428^{**}	0.941^{**}	0.589^{**}	1.055^{**}	0.267^{**}	0.826^{**}
	(0.013)	(0.037)	(0.018)	(0.056)	(0.015)	(0.053)
N	1620	1620	1620	1620	1620	1620
R^2	0.003	0.115	0.001	0.120	0.002	0.086

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

Table A2: Impact on Bad Outcomes (Red Balls Drawn) for the "Healthiest" (Urns with $\geq$ 90% Blue Balls) and "Sickest" (Urns with $\leq$ 10% Blue Balls)

	(1)	(2)	(3)	(4)	(5)	(6)
	% Red Balls for All Urns when One Urn is Healthiest or Sickest		% Red Balls for Sickest Urns		% Red Balls for Healthiest Urns	
Status Quo	0.072*	0.046+	0.130**	0.100*	0.074*	0.063+
	(0.032)	(0.027)	(0.045)	(0.044)	(0.036)	(0.036)
Jar # Blue Balls		-0.003**		-0.004**		-0.001
		(0.000)		(0.001)		(0.001)
High Urn # Blue Balls		-0.006**		-0.003**		-0.008
		(0.001)		(0.001)		(0.006)
Low Urn # Blue Balls		-0.003**		-0.014+		0.001
		(0.001)		(0.007)		(0.001)
Constant	0.411**	1.150**	0.731**	1.273**	0.062**	0.832
	(0.022)	(0.065)	(0.035)	(0.103)	(0.020)	(0.572)
N	539	539	300	300	295	295
R^2	0.012	0.255	0.026	0.129	0.016	0.038

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. This table reports only results from urns that are among the "Healthiest" (Urns with $\geq$ 90% Blue Balls) and "Sickest" (Urns with $\leq$ 10% Blue Balls. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

Table A3: Impact on Bad Outcomes (Red Balls Drawn) for those NOT "Healthiest" (Urns with $\geq$ 90% Blue Balls) or "Sickest" (Urns with $\leq$ 10% Blue Balls)

	(1)	(2)	(3)	(4)	(5)	(6)
	% Red Balls for all Urns		% Red Balls for Low Blue Urns		% Red Balls for High Blue Urns	
Status Quo	0.015	0.019	0.004	0.014	0.027	0.023
	(0.021)	(0.021)	(0.025)	(0.025)	(0.027)	(0.026)
Jar # Blue Balls		-0.001**		-0.003**		-0.001
		(0.000)		(0.000)		(0.000)
High Urn # Blue Balls		-0.003**		0.000		-0.007**
		(0.001)		(0.001)		(0.001)
Low Urn # Blue Balls		-0.003**		-0.008**		-0.000
		(0.000)		(0.001)		(0.001)
Constant	0.437**	0.856**	0.556**	1.012**	0.317**	0.819**
	(0.015)	(0.055)	(0.019)	(0.067)	(0.018)	(0.067)
N	1081	1081	1320	1320	1325	1325
R^2	0.000	0.058	0.000	0.086	0.001	0.049

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. This table reports only results from urns that are among those NOT "Healthiest" (Urns with $\geq$ 90% Blue Balls) or "Sickest" (Urns with $\leq$ 10% Blue Balls). The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

Table A4: Differences in Average Number of Blue Balls for Urns (Patients) and Jars (Kidneys) for Those Transplanted and Percentage Low Quality Jars Accepted

	(1)	(2)	(3)	(4)	(5)	(6)
	pre-TX Urn # Blue Balls		TXed Urn # Blue Balls	TXed Jar # Blue Balls		% Low Quality Jars Accepted | offered
Status Quo	13.47**	13.40**	6.990**	0.514	4.797**	-0.221**
	(1.567)	(1.553)	(1.277)	(1.836)	(1.135)	(0.0423)
Jar # Blue Balls		0.120**				
		(0.0329)				
Constant	41.81**	32.61**	59.07**	76.34**	78.87**	0.713**
	(1.016)	(2.854)	(0.791)	(1.053)	(0.896)	(0.0264)
Which transplants?	All	All	All	All	Did not made urn worse	Transplanted or not
N	1295	1295	1295	1295	1089	1094
R^2	0.064	0.073	0.039	0.000	0.023	0.050

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. The dependent variables are the number of blue balls of an urn (columns 1 and 2 are for the urn pre-transplantation, while column 3 is for the urn post-transplantation) or a jar (columns 4-5 are for the jars used in transplantation), and the percentage of Low Quality Jars accepted (column 6). The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in the jar.

Table A5: Impact on Bad Outcomes (Red Balls Drawn) for Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	% Red Balls for all Urns		% Red Balls for Low Blue Urns		% Red Balls for High Blue Urns	
Status Quo	0.034^{+}	0.031^{+}	0.026	0.031	0.041^{+}	0.032
	(0.018)	(0.0167)	(0.024)	(0.023)	(0.022)	(0.021)
Jar # Blue Balls		-0.002^{**}		-0.003^{**}		-0.001^{+}
		(0.000)		(0.000)		(0.000)
High Urn # Blue Balls		-0.004^{**}		-0.000		-0.007^{**}
		(0.000)		(0.001)		(0.001)
Low Urn # Blue Balls		-0.004^{**}		-0.008^{**}		0.000
		(0.000)		(0.001)		(0.001)
Constant	0.428^{**}	0.941^{**}	0.589^{**}	1.055^{**}	0.267^{**}	0.826^{**}
	(0.013)	(0.037)	(0.018)	(0.056)	(0.015)	(0.053)
N	1620	1620	1620	1620	1620	1620
R^2	0.003	0.115	0.001	0.120	0.002	0.086

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. The data here is from the alternative sample with revised experimental parameters. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

Table A6: Impact on Bad Outcomes (Red Balls Drawn) for the "Healthiest" (Urns with $\geq$ 90% Blue Balls) and "Sickest" (Urns with $\leq$ 10% Blue Balls) with Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	% Red Balls for All Urns when One Urn is Healthiest or Sickest		% Red Balls for Sickest Urns		% Red Balls for Healthiest Urns	
Status Quo	0.072*	0.046+	0.130**	0.100*	0.074*	0.063+
	(0.032)	(0.027)	(0.045)	(0.044)	(0.036)	(0.036)
Jar # Blue Balls		-0.003**		-0.004**		-0.001
		(0.000)		(0.001)		(0.001)
High Urn # Blue Balls		-0.006**		-0.003**		-0.008
		(0.001)		(0.001)		(0.006)
Low Urn # Blue Balls		-0.003**		-0.014+		0.001
		(0.001)		(0.007)		(0.001)
Constant	0.411**	1.150**	0.731**	1.273**	0.062**	0.832
	(0.022)	(0.065)	(0.035)	(0.103)	(0.020)	(0.572)
N	539	539	300	300	295	295
R^2	0.012	0.255	0.026	0.129	0.016	0.038

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. This table reports only results from urns that are among the "Healthiest" (Urns with $\geq$ 90% Blue Balls) and "Sickest" (Urns with $\leq$ 10% Blue Balls. The data here is from the alternative sample with revised experimental parameters. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

Table A7: Impact on Bad Outcomes (Red Balls Drawn) for those NOT "Healthiest" (Urns with $\geq$ 90% Blue Balls) or "Sickest" (Urns with $\leq$ 10% Blue Balls) with Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	% Red Balls for all Urns		% Red Balls for Low Blue Urns		% Red Balls for High Blue Urns	
Status Quo	0.015	0.019	0.004	0.014	0.027	0.023
	(0.021)	(0.021)	(0.025)	(0.025)	(0.027)	(0.026)
Jar # Blue Balls		-0.001**		-0.003**		-0.001
		(0.000)		(0.000)		(0.000)
High Urn # Blue Balls		-0.003**		0.000		-0.007**
		(0.001)		(0.001)		(0.001)
Low Urn # Blue Balls		-0.003**		-0.008**		-0.000
		(0.000)		(0.001)		(0.001)
Constant	0.437**	0.856**	0.556**	1.012**	0.317**	0.819**
	(0.015)	(0.055)	(0.019)	(0.067)	(0.018)	(0.067)
N	1081	1081	1320	1320	1325	1325
R^2	0.000	0.058	0.000	0.086	0.001	0.049

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. This table reports only results from urns that are among those NOT "Healthiest" (Urns with $\geq$ 90% Blue Balls) or "Sickest" (Urns with $\leq$ 10% Blue Balls). The data here is from the alternative sample with revised experimental parameters. The ball drawn from each urn at the end of each round could be either Blue or Red: We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage urns from which a red ball was drawn at the end of the round of that urn, representing the percentage of transplant candidates/patients getting a bad health outcome (e.g. mortality rate). The results here are conditional on Player 1 having recovered jars and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in each of the jars. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn.

Table A8: Differences in Average Number of Blue Balls for Urns (Patients) and Jars (Kidneys) for Those Transplanted and Percentage Low Quality Jars Accepted with Alternative Sample

	(1)	(2)	(3)	(4)	(5)	(6)
	pre-TX Urn # Blue Balls		TXed Urn # Blue Balls	TXed Jar # Blue Balls		% Low Quality Jars Accepted | offered
Status Quo	13.47**	13.40**	6.990**	0.514	4.797**	-0.221**
	(1.567)	(1.553)	(1.277)	(1.836)	(1.135)	(0.0423)
Jar # Blue Balls		0.120**				
		(0.0329)				
Constant	41.81**	32.61**	59.07**	76.34**	78.87**	0.713**
	(1.016)	(2.854)	(0.791)	(1.053)	(0.896)	(0.0264)
Which transplants?	All	All	All	All	Did not made urn worse	Transplanted or not
N	1295	1295	1295	1295	1089	1094
R^2	0.064	0.073	0.039	0.000	0.023	0.050

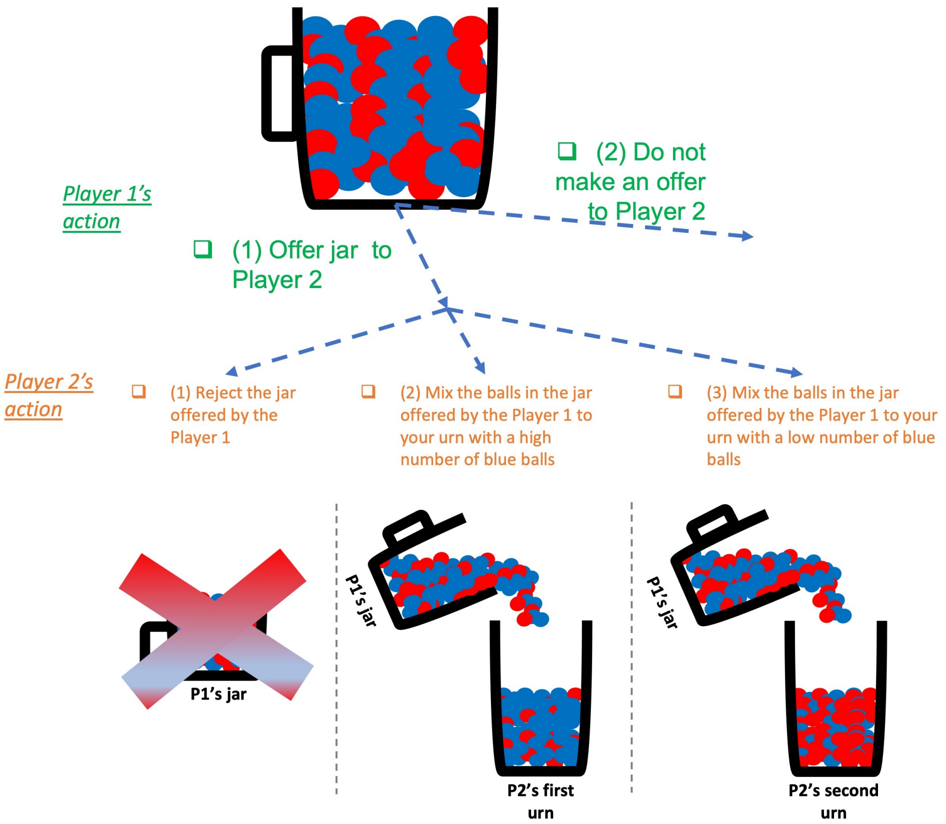
Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. The data here is from the alternative sample with revised experimental parameters. The dependent variables are the number of blue balls of an urn (columns 1 and 2 are for the urn pre-transplantation, while column 3 is for the urn post-transplantation) or a jar (columns 4-5 are for the jars used in transplantation), and the percentage of Low Quality Jars accepted (column 6). The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in the jar.

A.4 Pilots and Results from Pilots

Before the actual run of our experiment, between June 15, 2021 and July 13, 2021 we piloted variants of the main experiment in which the OPO is shown only 1 jar for each round. As the OPOs only get the opportunity to recover 1 jar per round, the TCs can at most conduct one mixing (and must leave the other urn as is). A simple graphic representing the stage game for each round in the pilot is shown in Figure A39 (note the difference between this and the analogous stage game illustrated by Figure I for the main experiment).

Figure A39: Stage Game for the OPO Player (P1) and TC Player (P2) in the Pilot Experiments



In total, 4 pilots were run. The first was run on June 16, 2021 (N=266), the second on June 25 (N=216), the third on July 6 (N=94), and the fourth on July 14 (N=172). The payoff/incentive parameters were largely similar to the main experiment reported in the body of the paper: the penalties for having all recovered jars declined (here 1 jar) and the rewards for having all recovered jars accepted (here 1 jar) for the OPO under the Status Quo condition, the penalty for drawing a red ball and reward for drawing a blue ball post-mixing for the TC under the Status Quo condition, and the penalty/reward for each red/blue ball drawn for both players under the Holistic condition are all the same as the main experiment. The difference between these 4 rounds is in the distributions (of red-versus-blue balls) from which the jars and urns are drawn in each round. These differences are illustrated in Figure A40.[44]

Insert Figure A40 about here.

The results, as represented by point estimates for the impact of the Status Quo treatment, on recovery, discards, and transplant rate are largely similar to the final experiment reported in this paper. The key difference is that with only 1 jar to use, the players cannot improve the final "health outcomes" as prominently as when there are 2 jars per round. Overall, within the 1-jar variants in the pilot stage, we varied the distributions of blue-versus-red balls across the jars and urns but the behavior remain largely similar in the signs of the effects.[45]

Figure A41 presents descriptive statistics that compares the key outcomes of the status quo and holistic conditions, and each of the subsequent sections will discuss the bar clusters shown in the Figure. Each panel in this Figure displays the results for a pilot.

[44] In the first and largest pilot, the distributions for the urns $U[40, 100]$ and $U[0, 60]$ for High and Low Blue urns respectively are exactly the same as in the experiment in the paper while the distribution for the jars are $U[80, 100]$ and $U[0, 80]$ for High and Low Quality jars respectively, instead of $U[70, 100]$ and $U[0, 70]$ in the main experiment. The distributions of jars and urns in pilot 1 are the most similar to the main experiment among the pilots.

In the other three pilots, we shift the distributions for jars and/or urns (e.g., in pilots 2 and 3 we change urns so that there are no "extreme" urns with almost all or no blue balls). In particular, in pilot 2 the distribution for the jars are $U[90, 100]$ and $U[50, 90]$ for High and Low Quality jars respectively while the distribution for the urns are $U[40, 100]$ and $U[20, 60]$ for High and Low Blue urns respectively. The key feature of pilot 2 is that jars of either type (Low and High) are helpful to improve the odds of drawing blue balls from either types of urns on average (while Low Quality jars are not helpful for High Blue urns on average in the main experiment and in the other pilots).

In pilot 3, the distribution for the jars are $U[10, 50]$ and $U[0, 10]$ for High and Low Quality jars respectively while the distribution for the urns are $U[40, 100]$ and $U[20, 60]$ for High and Low Blue urns respectively. Pilot 3 is the opposite of pilot 2 in that jars of both types are not helpful to improve the odds of drawing blue balls from either types of urns on average.

In pilot 4, we increased the support of the distribution of the "High Blue" urn and reduced the support of the distribution of the "Low Blue" urn relative to pilot 1 (and the main experiment), and shifted the distribution for jars slightly. In particular, the distribution for the jars are $U[60, 100]$ and $U[0, 60]$ for High and Low Quality jars respectively while the distribution for the urns are $U[20, 100]$ and $U[0, 25]$ for High and Low Blue urns respectively. In pilot 4, the opportunity to improve the odds of drawing blue balls for High Blue urns through mixing is increased. Also, note that Low Blue urns can be improved by any mixing.

[45] The only exception is the effect of Status Quo on "Bad Discards" for Pilot 2 (see Table A10).

Insert Figure A41 about here.

A.4.1 Comparing Recovery, Discards, Transplants Between Pilots and the Main Experiment

The first pair of bars of for each Panel in Figure A41 displays the average rates of jar recovery and offering by Player 1. In the main experiment, Player 1s recover 17.4 percentage points fewer jars under status quo compared to holistic regulations: 58.6% and 76.0% of the jars are recovered under status quo and holistic conditions respectively. Across every one of the pilots, Player 1s recovers significantly fewer jars under status quo compared to holistic regulations: 17.6, 15.6, 15.5, and 25.1 percentage points fewer jars for pilots 1, 2, 3, and 4 respectively (see column (1) of Table A9).

In the main experiment, there is no statistically significant[46] difference in recovery rates when the jars are high quality jars. The same is true for each of the four pilots (see column (3) of Table A9). Similar to the main experiment, Player 1s recover similarly when the observed jar quality type is high (number of blue balls), but the recovery behavior diverges when the observed jar quality type is low in the pilots as well.

Insert Table A9 about here.

The second pair of bars of Figure A41 and columns 4 to 6 of Table A9 show the results on instances where a pair of jars could have benefited at least one urn but was not recovered. In the main experiment, we reported that under the status quo condition, Player 1 "missed recovery opportunities" 26.2% of the time while missed recovery is only 15.7% under the holistic condition. The differences are significantly higher by 10.5 percentage points. The sign of the differences in "missed recovery opportunities" from the four pilots are all similar to the main experiment with three of them significantly so (the difference is not significant for pilot 3 where, by design, Low Quality jars can never be a "missed recovery opportunity" because they always have fewer blue balls than any urns that can be drawn.).

The third pair of bars of Figure A41 and the first column of Table A10 show that the average quality (average number of blue balls) of recovered and offered jars is higher

[46] At conventional levels.

under the status quo condition for the main experiment as well as for all four pilots. The average recovered jar has a statistically significant 6.0 more blue balls (out of 100) under the holistic condition for the main experiment. Likewise, the average recovered jar has statistically significantly (5%-level) more blue balls (out of 100) for pilots 1, 2, and 4 (ranging from 2.9 to 11.2 more blue balls). We also see a higher average number of blue balls of recovered and offered jars (3.5 more blue balls) for pilot 3 but due to the small sample size (less than half of the next smallest pilot sample size), we did not obtain a coefficient that is significant at conventional levels.

The evidence for cherry picking done by Player 1s is present in the pilots as it is for the main experiment.

Insert Table A10 about here.

The fourth pair of bars of Figure A41 and columns 2 to 4 of Table A10 display the results on jar discards by Player 2 when Player 1 offered her a jar. In the main experiment, the lower average quality of offered jars did not lead to significantly more discards under the holistic regulations condition. Across the four pilots, there are either no significant difference in discards or significantly lower discards under the holistic regulations condition. This means that, in both the main experiment and the pilots, the threshold for accepting a jar to carry out a transplant is lower under the holistic regulations condition.

The fifth pair of bars of Figure A41 and columns 5 to 7 in Table A10 show that Player 2s (TC) discard jars that could have helped the blue-ball-odds of an urn in a level that is 23.8 percentage points more under status quo condition than under holistic regulations in the main experiment. For this outcome variable, we do see some divergence from the pilot results from the main experiment. While the point estimate for these "bad discards" are similar between that from pilots 1, 3, and 4, and the main experiment, none of these point estimates from these pilots are significant at conventional levels. In pilot 2, the sign of the point estimate is actually the opposite of the estimate from the main experiment.

The sixth pair of bars of Figure A41 and Table A11 displays the results on mixing ("transplants") by Player 2. On average, Player 2s mix a significant 8.8 percentage points less under status quo compared to holistic regulations in the main experiment. Again, the analogous point estimates from the pilots are of the same sign as that from the main experiment. These differences in mixing rate are statistically significantly (5%-level) for pilots 1, 2, and 4 (ranging from 12.2 to 26.2 percentage points less for Status Quo). We also see a lower mixing rate for Status Quo (2.9 percentage points) for pilot 3 but due to the small sample size, we did not obtain a coefficient that is significant at conventional levels.

Insert Table A11 about here.

The seventh pair of bars of Figure A41 and columns 4 to 6 in Table A11 show that Player 2s conduct significantly many (14.6%) more harmful transplants under the status quo condition than under holistic regulations in the main experiment and significantly many more harmful transplants under the status quo in pilots 1, 3, and 4. These results are significant at conventional levels, controlling for jar/urn quality or not. While the point estimate, similarly signed as the main experiment, is not significant pilot 2 when not controlling for jar/urn quality, it becomes significant once we control for just jar or both jar and urn qualities.

A.4.2 Comparing "Health Outcomes" Between Pilots and the Main Experiment

The last pair of bars of Figure A41 and Table A12 show that the expected number of red balls drawn is 4.1 percentage points higher under Status Quo for the main experiment. The analogous point estimates from the pilots are of the same sign as that from the main experiment. These differences in expected number of red balls are statistically significantly (1%-level) for pilots 1, 2, and 4 (ranging from 3.1 to 4.4 percentage points higher for Status Quo). The expected number of red balls is also higher for Status Quo (0.8 percentage points) for pilot 3 but due to the small sample size, we did not obtain a coefficient that is significant at conventional levels.

It is worth noting that part of the reason why we varied the distributions of jars and urns during the pilot was to investigate the impact of the actual number of red balls drawn (as opposed to expected number of red balls drawn, which is the more reasonable proxy for expected health outcome impact, as we argued in Section 4.3).

Insert Table A12 about here.

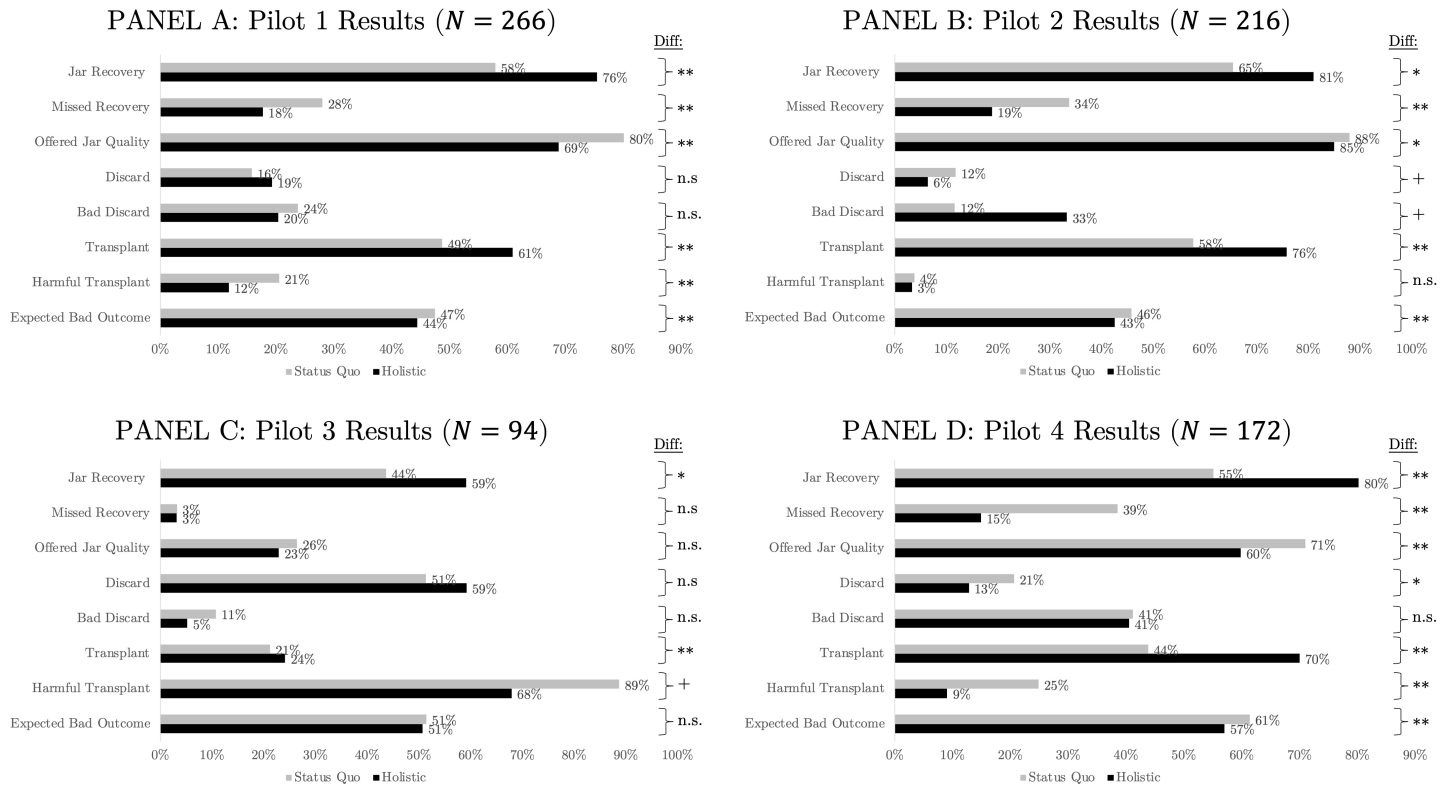
The results from the main experiment on missed recoveries, bad discards, and bad mixings (transplants) offer evidence for efficiency loss. We can find support for similar efficiency loss from the pilots, despite their different game structure (with only one jar per round and different distributions for jars/urns). The results from the pilots lend additional support for the takeaway that holistic regulations might benefit through incentivizing better behavior and consequently better health outcomes.

Figure A40: Comparison between the Main Experiment with the Pilot Experiments

	# jars per round	Status Quo Payoffs	Holistic Payoffs	Distribution (Blue Balls): High Quality Jar	Low Quality Jar	High Blue Urn	Low Blue Urn
Main Experiment	• 2 jars	• P1: +$0.1 for 2 jars; $0.0 for 1 jar; -$0.3 for 0 jars (accepted if offered) • P2: +$0.25/blue; -$1.00/red (if mix)	• P1: +$0.16/blue; -$0.08 /red • P2: +$0.20/blue; -$0.10 /red	• U[70,100]	• U[0,70]	• U[40,100]	• U[0,60]
Pilot 1				• U[80,100]	• U[0,80]	• U[40,100]	• U[0,60]
Pilot 2	• 1 jar	• P1: +$0.1 for 1 jar; -$0.3 for 0 jars (accepted if offered) • P2: +$0.25/blue; -$1.00/red (if mix)	• P1: +$0.16/blue; -$0.08 /red • P2: +$0.20/blue; -$0.10 /red	• U[90,100]	• U[50,90]	• U[40,90]	• U[20,60]
Pilot 3				• U[10,50]	• U[0,10]	• U[40,90]	• U[20,60]
Pilot 4				• U[60,100]	• U[0,60]	• U[20,100]	• U[0,25]
Main Experiment (Alternative parameters)	• 2 jars	• P1: $0.1 for 2 jars; $0.0 for 1 jar; -$0.3 for 0 jars (accepted if offered) • P2: +$0.11175/blue; -$0.1 /red (if mix)	• P1: +$0.0957/blue; -$0.0957 /red • P2: +$0.1/blue; -$0.1 /red	• U[70,100]	• U[0,70]	• U[40,100]	• U[0,60]

Notes: This table compares the Main Experiment with the pilots.

Figure A41: Pilot Outcome Variable Comparison



Standard errors clustered by player-pairs $^{+}$ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$

Notes: This Figure shows averages ("mean") by treatment condition. The significance-levels of the differences between treatment conditions are reported. "Jar Recovery" is the percentage of the pairs of jars that are recovered and offered by Player 1. "Missed Recovery" is the percentage of the pairs of jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1. "Offered Jar Quality" is the percentage of balls that are blue in each of the recovered and offered jars, "Discard" is the percentage of recovered jars that are discarded/rejected. "Bad Discard" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn). "Transplant" is the percentage of urns where a mixing happened. "Harmful Transplant" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn. "Expected Bad Outcome" is the proportion of red balls in an urn at the end of a round.

Table A9: Impact on Jar Recovery Rates and Missed Opportunities for Beneficial Recoveries

	(1)	(2)	(3)	(4)	(5)	(6)
		Jar Recovery Rate			Missed Beneficial Recovery	
	All	Low Quality	High Quality	All	Low Quality	High Quality
			Pilot 1			
Status Quo	-0.176**	-0.368**	0.004	0.103**	0.216**	-0.004
	(0.039)	(0.064)	(0.044)	(0.033)	(0.051)	(0.044)
Constant	0.755**	0.610**	0.904**	0.177**	0.255**	0.096**
	(0.032)	(0.053)	(0.032)	(0.027)	(0.039)	(0.032)
N	1330	668	662	1330	668	662
R^2	0.035	0.139	0.000	0.015	0.050	0.000
			Pilot 2			
Status Quo	-0.156**	-0.315**	0.014	0.149**	0.302**	-0.014
	(0.043)	(0.073)	(0.039)	(0.043)	(0.073)	(0.039)
Constant	0.810**	0.708**	0.906**	0.188**	0.289**	0.0936**
	(0.033)	(0.054)	(0.032)	(0.033)	(0.054)	(0.032)
N	1080	536	544	1080	536	544
R^2	0.031	0.100	0.001	0.028	0.092	0.001
			Pilot 3			
Status Quo	-0.155*	-0.229*	-0.009	0.000		0.00870
	(0.058)	(0.095)	(0.091)	(0.023)		(0.044)
Constant	0.591**	0.362**	0.800**	0.0318		0.0609
	(0.042)	(0.080)	(0.072)	(0.020)		(0.037)
N	470	240	230	470		230
R^2	0.024	0.072	0.000	0.000		0.000
			Pilot 4			
Status Quo	-0.251**	-0.446**	-0.089	0.236**	0.408**	0.089
	(0.042)	(0.068)	(0.056)	(0.039)	(0.058)	(0.056)
Constant	0.802**	0.658**	0.942**	0.149**	0.243**	0.058*
	(0.029)	(0.053)	(0.028)	(0.025)	(0.040)	(0.028)
N	860	414	446	860	414	446
R^2	0.071	0.203	0.021	0.070	0.168	0.021
			Main Experiment			
Status Quo	-0.174**	-0.299**	-0.042	0.105**	0.166**	0.042
	(0.035)	(0.055)	(0.038)	(0.026)	(0.036)	(0.038)
Constant	0.760**	0.610**	0.913**	0.157**	0.225**	0.087**
	(0.024)	(0.041)	(0.022)	(0.017)	(0.024)	(0.022)
N	1620	820	800	1620	820	800
R^2	0.035	0.090	0.004	0.017	0.032	0.004

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + \epsilon_i$. Player 1 (OPO) can either recover a jar or not (1=recover/offer jar; 0=not recover/offer jar). The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The 1^{st} and 4^{th} columns reports results unconditional on jar quality type, the 2^{nd} and 5^{th} columns reports results conditional on the jar quality being low quality, and the 3^{rd} and 6^{th} columns reports results conditional on the jar quality being high quality. "Jar Recovery Rate" is the percentage of the jars that are recovered and offered by Player 1. "Missed Beneficial Recovery" is the percentage of the jars that could have improved the odds of a good outcome (drawing a blue ball) of at least one urn but are NOT recovered/offered by Player 1. Finally, note that for Pilot 3, we do not have any missed beneficial recoveries because the Low Quality jars can only have at most 10 blue balls while all the urns have at least 20 blue balls - so the Low Quality jars cannot have improved the odds of drawing a blue ball for any of the urns given the blue ball distribution in Pilot 3.

Table A10: Impact on Recovered Jar Quality, Discard Rates, and Discards that could have Benefited Urn(s)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Jar # Blue | Recovered	Jar Discard Rate | Recovered			% of Bad Discards		
				Pilot 1			
Status Quo	11.220**	-0.035	0.048^{+}	0.050^{+}	0.034	0.017	0.018
	(2.251)	(0.0319)	(0.0290)	(0.0283)	(0.0656)	(0.0706)	(0.0669)
Jar # Blue Balls			-0.007**	-0.007**		0.001	0.001
			(0.001)	(0.001)		(0.001)	(0.001)
High Urn # Blue Balls				-0.000			0.006**
				(0.001)			(0.002)
Low Urn # Blue Balls				0.002**			-0.007**
				(0.001)			(0.002)
Constant	68.870**	0.193**	0.702**	0.659**	0.204**	0.172**	-0.020
	(1.717)	(0.022)	(0.055)	(0.071)	(0.041)	(0.045)	(0.170)
N	882	882	882	882	156	156	156
R^2	0.043	0.002	0.262	0.275	0.002	0.006	0.168
				Pilot 2			
Status Quo	2.920*	0.053^{+}	0.070*	0.069*	-0.217^{+}	-0.148	-0.035
	(1.148)	(0.028)	(0.029)	(0.028)	(0.114)	(0.094)	(0.093)
Jar # Blue Balls			-0.006**	-0.006**		-0.011**	-0.010**
			(0.001)	(0.001)		(0.003)	(0.003)
High Urn # Blue Balls				-0.002*			0.011**
				(0.001)			(0.004)
Low Urn # Blue Balls				-0.001			-0.006
				(0.001)			(0.003)
Constant	85.000**	0.064**	0.559**	0.737**	0.333**	1.122**	0.603^{+}
	(0.765)	(0.020)	(0.094)	(0.128)	(0.102)	(0.236)	(0.305)
N	787	787	787	787	70	70	70
R^2	0.011	0.009	0.086	0.096	0.070	0.285	0.382
				Pilot 3			
Status Quo	3.517	-0.079	-0.041	-0.049	0.055	0.0486	0.047
	(2.537)	(0.091)	(0.087)	(0.087)	(0.054)	(0.053)	(0.051)
Jar # Blue Balls			-0.011**	-0.011**		0.005*	0.007**
			(0.002)	(0.002)		(0.002)	(0.002)
High Urn # Blue Balls				-0.003			0.001
				(0.003)			(0.002)
Low Urn # Blue Balls				0.004			-0.009**
				(0.003)			(0.003)
Constant	22.920**	0.592**	0.840**	0.837**	0.052^{+}	-0.046	0.247^{+}
	(1.795)	(0.060)	(0.085)	(0.182)	(0.030)	(0.030)	(0.132)
N	239	239	239	239	133	133	133
R^2	0.013	0.006	0.119	0.135	0.011	0.086	0.246
				Pilot 4			
Status Quo	11.150**	0.078*	0.163**	0.166**	0.007	0.048	0.110
	(2.351)	(0.034)	(0.029)	(0.029)	(0.108)	(0.130)	(0.135)
Jar # Blue Balls			-0.008**	-0.008**		-0.002	-0.001
			(0.001)	(0.001)		(0.002)	(0.002)
High Urn # Blue Balls				-0.002*			0.008**
				(0.001)			(0.002)
Low Urn # Blue Balls				-0.001			-0.011
				(0.002)			(0.007)
Constant	59.820**	0.128**	0.584**	0.699**	0.405**	0.436**	0.055
	(1.543)	(0.022)	(0.054)	(0.079)	(0.073)	(0.087)	(0.171)
N	577	577	577	577	93	93	93
R^2	0.044	0.011	0.297	0.309	0.000	0.006	0.158
				Main Experiment			
Status Quo	5.985**	-0.022	0.020	0.024	0.238**	0.143**	0.060*
	(2.106)	(0.029)	(0.027)	(0.027)	(0.046)	(0.039)	(0.030)
Jar # Blue Balls			-0.007**	-0.007**		0.008**	0.010**
			(0.000)	(0.000)		(0.001)	(0.000)
High Urn # Blue Balls				0.003**			-0.013**
				(0.001)			(0.001)
Low Urn # Blue Balls				0.000			-0.002*
				(0.001)			(0.001)
Constant	65.060**	0.418**	0.878**	0.686**	0.260**	-0.191**	0.777**
	(1.379)	(0.021)	(0.030)	(0.054)	(0.032)	(0.021)	(0.053)
N	1094	1094	1094	1094	702	702	702
R^2	0.012	0.001	0.300	0.317	0.064	0.323	0.564

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either discard 0 or 1 jar *if a jar was offered by Player 1*. The results here are conditional on Player 1 having recovered a jar and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in the jar. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Jar # Blue | Recovered" is the number of blue balls in the recovered and offered jar (out of 100), "Jar Discard Rate | Recovered" is the percentage of recovered jars that are discarded/rejected. "% of Bad Discards" is the percentage of discarded jars that has more blue balls than at least one urn (benefits at least one urn).

Table A11: Impact on Mixing ("Transplant") Rate, and Mixings that Gave an Urn Worse Odds for Blue

	(1)	(2)	(3)	(4)	(5)	(6)
	Mixing (Transplant) Rate			% Mixings Made Urn Worse		
Pilot 1						
Status Quo	-0.122**	-0.139**	-0.140**	0.087**	0.147**	0.146**
	(0.035)	(0.034)	(0.034)	(0.031)	(0.026)	(0.025)
Jar # Blue Balls		0.009**	0.009**		-0.007**	-0.008**
		(0.001)	(0.001)		(0.001)	(0.001)
High Urn # Blue Balls			0.000			0.007**
			(0.001)			(0.001)
Low Urn # Blue Balls			-0.001+			0.002**
			(0.001)			(0.001)
Constant	0.609**	0.014	0.044	0.118**	0.685**	0.156+
	(0.027)	(0.042)	(0.070)	(0.023)	(0.087)	(0.089)
N	1330	1330	1330	726	726	726
R^2	0.015	0.326	0.328	0.014	0.187	0.310
Pilot 2						
Status Quo	-0.181**	-0.178**	-0.175**	0.004	0.025+	0.024+
	(0.047)	(0.044)	(0.044)	(0.015)	(0.014)	(0.014)
Jar # Blue Balls		0.013**	0.013**		-0.006**	-0.006**
		(0.001)	(0.001)		(0.001)	(0.001)
High Urn # Blue Balls			0.003**			0.002**
			(0.001)			(0.000)
Low Urn # Blue Balls			0.000			0.001*
			(0.001)			(0.001)
Constant	0.758**	-0.328**	-0.525**	0.0330**	0.557**	0.409**
	(0.036)	(0.117)	(0.143)	(0.010)	(0.109)	(0.092)
N	1080	1080	1080	717	717	717
R^2	0.037	0.199	0.205	0.000	0.177	0.192
Pilot 3						
Status Quo	-0.029	-0.017	-0.019	0.208+	0.263**	0.234*
	(0.054)	(0.052)	(0.052)	(0.103)	(0.091)	(0.089)
Jar # Blue Balls		0.012**	0.012**		-0.013**	-0.012**
		(0.001)	(0.001)		(0.002)	(0.002)
High Urn # Blue Balls			0.001			0.003
			(0.002)			(0.003)
Low Urn # Blue Balls			-0.002			0.007*
			(0.002)			(0.003)
Constant	0.241**	0.022	0.065	0.679**	1.035**	0.557*
	(0.039)	(0.042)	(0.110)	(0.093)	(0.060)	(0.245)
N	470	470	470	106	106	106
R^2	0.001	0.201	0.207	0.063	0.273	0.313
Pilot 4						
Status Quo	-0.262**	-0.287**	-0.287**	0.158**	0.217**	0.179**
	(0.043)	(0.041)	(0.041)	(0.041)	(0.041)	(0.038)
Jar # Blue Balls		0.010**	0.010**		-0.005**	-0.005**
		(0.001)	(0.001)		(0.001)	(0.001)
High Urn # Blue Balls			0.001*			0.006**
			(0.001)			(0.001)
Low Urn # Blue Balls			0.002			0.002
			(0.002)			(0.002)
Constant	0.700**	0.158**	0.069	0.091**	0.421**	0.029
	(0.031)	(0.050)	(0.064)	(0.020)	(0.070)	(0.077)
N	860	860	860	484	484	484
R^2	0.070	0.388	0.391	0.046	0.131	0.271
Main Experiment						
Status Quo	-0.088**	-0.087**	-0.088**	0.146**	0.166**	0.163**
	(0.028)	(0.025)	(0.025)	(0.030)	(0.023)	(0.023)
Jar # Blue Balls		0.008**	0.008**		-0.008**	-0.008**
		(0.000)	(0.000)		(0.001)	(0.001)
High Urn # Blue Balls			-0.001**			0.004**
			(0.001)			(0.001)
Low Urn # Blue Balls			-0.000			0.002**
			(0.000)			(0.001)
Constant	0.443**	-0.018	0.086+	0.083**	0.685**	0.331**
	(0.020)	(0.023)	(0.047)	(0.016)	(0.056)	(0.062)
N	1620	1620	1620	903	903	903
R^2	0.012	0.332	0.336	0.042	0.297	0.352

Standard errors (Robust, clustered by player-pairings) in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + X_i\gamma + \epsilon_i$. Player 2 (TC) can either mix 0 or 1 jar into the urns *if a jar was offered by Player 1.* The results here are conditional on Player 1 having recovered a jar and made an offer to Player 2. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. "Jar # Blue Balls" is the number of blue balls in the jar. "High Urn # Blue Balls" is the number of blue balls in the high blue urn. "Low Urn # Blue Balls" is the number of blue balls in the low blue urn. "Mixing (Transplant) Rate" is the percentage of urns where a mixing happened, this is the number of transplant(s) divided by 1 per round. "% Mixings Made Urn Worse" is the percentage of mixings that occurred which made the odds of blue balls worse for an urn.

Table A12: Impact on Expected Bad Outcomes (Red Balls Drawn) Based on Actual Mixing Behavior

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All Urns		At Least One Urn Sickest or Healthiest			No Urn Sickest or Healthiest		
	All	Low Blue	High Blue	All	Low Blue	High Blue	All	Low Blue	High Blue
				Pilot 1					
Status Quo	0.031^{**}	0.076^{**}	-0.015^{+}	0.061^{**}	0.118^{**}	0.004	0.016^{*}	0.051^{**}	-0.019^{*}
	(0.008)	(0.013)	(0.009)	(0.015)	(0.021)	(0.016)	(0.007)	(0.012)	(0.010)
Constant	0.444^{**}	0.601^{**}	0.286^{**}	0.425^{**}	0.675^{**}	0.176^{**}	0.452^{**}	0.567^{**}	0.337^{**}
	(0.006)	(0.010)	(0.006)	(0.012)	(0.016)	(0.012)	(0.006)	(0.010)	(0.009)
N	2660	1330	1330	878	439	439	1782	891	891
R^2	0.004	0.038	0.002	0.008	0.078	0.000	0.001	0.023	0.005
				Pilot 2					
Status Quo	0.032^{**}	0.070^{**}	-0.005				0.032^{**}	0.070^{**}	-0.005
	(0.007)	(0.013)	(0.009)				(0.007)	(0.013)	(0.009)
Constant	0.425^{**}	0.499^{**}	0.350^{**}				0.425^{**}	0.499^{**}	0.350^{**}
	(0.006)	(0.011)	(0.006)				(0.006)	(0.011)	(0.006)
N	2160	1080	1080				2160	1080	1080
R^2	0.009	0.052	0.000				0.009	0.052	0.000
				Pilot 3					
Status Quo	0.007	0.002	0.012				0.007	0.002	0.012
	(0.008)	(0.008)	(0.013)				(0.008)	(0.008)	(0.013)
Constant	0.507^{**}	0.600^{**}	0.414^{**}				0.507^{**}	0.600^{**}	0.414^{**}
	(0.005)	(0.006)	(0.011)				(0.005)	(0.007)	(0.011)
N	940	470	470				940	470	470
R^2	0.001	0.000	0.002				0.001	0.000	0.002
				Pilot 4					
Status Quo	0.044^{**}	0.097^{**}	-0.009	0.027^{+}	0.010^{**}	-0.045^{+}	0.062^{**}	0.087^{**}	0.037^{*}
	(0.009)	(0.013)	(0.014)	(0.014)	(0.017)	(0.023)	(0.011)	(0.013)	(0.016)
Constant	0.570^{**}	0.757^{**}	0.383^{**}	0.576^{**}	0.798^{**}	0.354^{**}	0.564^{**}	0.716^{**}	0.412^{**}
	(0.007)	(0.011)	(0.009)	(0.010)	(0.016)	(0.015)	(0.009)	(0.011)	(0.012)
N	1720	860	860	904	452	452	816	408	408
R^2	0.006	0.109	0.000	0.002	0.107	0.010	0.019	0.124	0.010
					Main Experiment				
Status Quo	0.041^{**}	0.055^{**}	0.027^{**}	0.070^{**}	0.101^{**}	0.039^{*}	0.026^{**}	0.034^{**}	0.018^{*}
	(0.007)	(0.012)	(0.008)	(0.016)	(0.022)	(0.015)	(0.007)	(0.011)	(0.009)
Constant	0.421^{**}	0.578^{**}	0.264^{**}	0.401^{**}	0.638^{**}	0.164^{**}	0.432^{**}	0.548^{**}	0.315^{**}
	(0.006)	(0.008)	(0.006)	(0.011)	(0.016)	(0.010)	(0.006)	(0.008)	(0.007)
N	3240	1620	1620	1078	539	539	2162	1081	1081
R^2	0.007	0.020	0.007	0.012	0.054	0.015	0.004	0.010	0.004

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table presents the estimated parameter results for the estimation model $Y_i = \beta_0 + \beta_1 StatusQuo_i + \epsilon_i$. We model bad health outcomes as the drawing of red balls. The outcome of interest reported in this table is "% Red Ball." This outcome is the percentage of balls that are red balls in urns at the end of the round, representing the percentage of transplant candidates/patients expected to get a bad health outcome (e.g. expected mortality rate). The results here are conditional on actual Player 1 and Player 2 behavior in the game but not the actual draws. The independent variable "Status Quo" indicates the status quo condition where the incentives resemble the current fragmented regulations. The dependent variable in this table is the percentage of balls in an urn that is red at the end of a round. Note that since the urns in Pilots 2 and 3 cannot be either sickest (less than 10 blue balls) or healthiest (more than 90 blue balls) since the urns in these pilots can at most have 90 blue balls and at least have 20 blue balls. As a result the results for columns (4) to (6) are N/A for Pilots 2 and 3.