ON THE EMERGENCE OF MINORITY DISADVANTAGE: TESTING THE CULTURAL RED KING HYPOTHESIS

AYDIN MOHSENI, CAILIN O'CONNOR, AND HANNAH RUBIN

Abstract. The study of social justice asks: what sorts of social arrangements are equitable ones? But also: how do we derive the inequitable arrangements we often observe in human societies? In particular, in spite of explicitly stated equity norms, categorical inequity tends to be the rule rather than the exception. The cultural Red King hypothesis predicts that differentials in group size may lead to inequitable outcomes for minority groups even in the absence of explicit or implicit bias. We test this prediction in an experimental context where subjects divided into groups engage in repeated play of a bargaining game. We ran 14 trials involving a total of 112 participants. The results of the experiments are statistically significant and suggestive: individuals in minority groups in these experiments end up receiving fewer resources than those in majority groups. Combined with previous theoretical findings, these results give some reason to think that the cultural Red King may occur in real human groups.

1. Introduction

The study of social justice asks: what sorts of social arrangements are equitable ones? But also: how do we derive the inequitable arrangements we often observe in human societies? In particular, often in spite of explicitly stated equity norms, categorical inequity tends to be the rule rather than the exception.[1] Wherever humans recognize social categories—gender, race, religion etc.—status inequities, power inequities, and economic inequities tend to emerge across these divisions. There are many reasons such inequities emerge. In this paper we devise an experiment to test one mechanism that has been proposed to explain the emergence of minority disadvantage in particular—the cultural Red King effect.

This effect was first described by political philosopher Justin Bruner, who uses evolutionary game theoretic methods to show how minority groups can be disadvantaged in the emergence of bargaining conventions solely by dint of their group

[1] Adams and Freedman (1976); Freedman and Montanari (1980) discuss the prevalence of equity norms. Tilly et al. (1998) describes research on the prevalence of categorical inequity.

size (Bruner 2017). As he shows, in groups with completely symmetric preferences, abilities, and resources, minority status alone can increase the likelihood that individuals end up with fewer economic resources. The driver behind this effect is a learning asymmetry between minority and majority groups. While minority members commonly meet their out-group, the reverse is not true. As a result, members of a minority will more quickly learn to interact with their out-group.[2] In situations where this learning is about bargaining interactions, this often proves disadvantageous. Low, accommodating demands tend to be more safe in bargaining interactions, meaning that swift learners should adopt these demands. Once this is done, members of the majority group can take advantage of this accommodation.

Subsequent work has shown that this effect arises robustly in cultural evolutionary models (O'Connor 2017; O'Connor and Bruner 2017). Given the simplicity of these models, though, a further question arises: can the cultural Red King really occur in human groups? If so, there are important consequences for political philosophy. To give one example, consider accounts of social justice that appeal to historical justice in the sense outlined by Nozick (1974). The general idea is that distributions of wealth derived from just processes are just.[3] The models mentioned represent individuals who gain access to goods by willingly entering into bargaining agreements, in doing so employing strategies that best benefit them given their social arrangement. We might well want to describe the interactions they engage in as just ones. And yet, under these conditions entire classes of people end up disadvantaged for no reason besides their minority status.[4]

Of course, as noted, highly simplified models of social interaction cannot usually be taken at face value as explaining real social phenomena. One important epistemic role they can play is directing attention to processes that might be occurring in the real world, and which merit further empirical investigation. For this reason, we study the cultural Red King effect in the laboratory. In particular, we draw on tools from experimental economics. These allow us to create an environment where actors in groups bargain for real money, and where the

[2] There is some relation here to inversion thesis from standpoint epistemology. The idea there is that sometimes those in oppressed groups have special access to knowledge or epistemic privilege (Wylie 2013). In Bruner's models, minority groups learn more quickly about their social situation.

[3] Nozick (1974) in particular requires that such distributions derive from just initial acquisitions of goods and just transfers of holdings, though he does not do much fill in the details of what justice here entails.

[4] Many thanks to Nick Adams and Mike Ashfield for this point.

only asymmetry between them is group size. In this way, we are able to control conditions so that if we systematically see an advantage arising for large groups, we can conclude that the cultural Red King can potentially occur among human actors.

In this study, we found support for the cultural Red King effect. Members of minority groups ended up earning less money than those in majority groups on average. And this difference emerged over the course of an experiment where individuals learned to bargain. Further investigation will be important in confirming the significance of this effect to the real world. But given the theoretical work underpinning the result, there is reason to think that minority status alone can confer bargaining disadvantage to real world groups.

The paper will proceed as follows. In section 2 we introduce some preliminaries necessary to describe both the theoretical and experimental results on the cultural Red King that we discuss in this paper. In particular, we give a brief overview of the game theoretic and evolutionary game theoretic methods used to first describe the cultural Red King, followed by an overview of the basic tenets of experimental economics. In section 3, we move on to describe the cultural Red King effect, and the conditions under which it is expected to arise. We also state our experimental predictions more fully. Section 4 describes our experimental setup, and in 5 we give our findings. In section 6 we conclude.

2. Preliminaries: Game Theory, Evolutionary Game Theory, and Experimental Economics

The cultural evolutionary pathways that lead to the emergence of norms are notoriously difficult to investigate empirically. In particular, norms emerge among many individuals over long time scales. As such, many have turned to formal models to illuminate such pathways. Evolutionary game theory in particular has been useful in showing how certain norms or conventions might have emerged.

Game theoretic models start with a simplified representation of some strategic situation of interest. These *games* include the *players* involved in the strategic interaction, the possible *strategies* they might employ, and *payoffs* which represent how much each player values each of the outcomes. Game theorists generally explain outcomes or predict which outcomes to expect based on what rational decision makers would choose. For example, the concept of a *Nash equilibrium*, a combination of strategies where no actor can gain a better payoff by unilaterally changing their strategy, is often employed to explain strategic behavior.

Evolutionary game theory adds to these elements a *dynamics*, or description of how strategies might change over time, allowing evolutionary game theorists to explain the emergence of cultural norms as the result of some sort of learning process. Evolutionary game theorists generally explain outcomes or predict which outcomes to expect based on how likely they are to emerge as stable endpoints, or *equilibria*, of various learning processes. As will become clear in the next section, this framework has been very useful to modelers interested in explaining categorical inequity.

While experiments in philosophy often make use of methods from psychology, here we use methods of experimental economics, which are well-suited to test predictions from evolutionary game theoretic models. The cornerstone of experimental economics is *induced valuation*, where subjects' payment is determined by their (and, generally, also other subjects') decision making during the experiment (Smith 1976). The idea behind induced valuation is to ensure that subjects care about the outcomes of the strategic interaction in the laboratory and make real decisions in pursuit of outcomes. In this way, experimenters can draw conclusions about what subjects actually do in certain situations (rather that what subjects report they will do) (Croson 2005).

Another important feature of experimental economics is that these experiments tend to be, by and large, *context-free*. That is, the strategic situations are presented to subjects in a way that is as abstract as possible, without framing or reference to real world examples. This is done in order to avoid subjects bringing in preconceived notions about what the outcomes should be. So, for instance, our experiment is not framed in terms of investigating discrimination or inequity: this could make subjects less likely to request high amounts of resource as a result of cultural norms or ethical concerns, rather than being motivated purely by the payoffs of the game.

A final feature of experimental economics is that, as a rule, experimenters do not deceive their subjects. This *no deception* rule is important because subjects have to believe the strategic structure presented to them is really the structure they are making decisions within. For instance, subjects often believe that experimenters actively try to minimize what they pay their subjects (Cooper 2014). Subjects in our experiment must believe that they are interacting with other subjects, rather than thinking the experimenter is manipulating outcomes, so that the norm that evolves reflects real learning in response to the behavior of others.

3. Theory and Predictions

We will proceed by first describing the general theory behind the cultural Red King effect. Then we will turn to the more specific theory underlying our experimental design. We will use this to outline our experimental predictions.

3.1. **Theory.** According to the Red Queen hypothesis in biology, fast evolving species can gain an advantage over slow evolving ones. In a predator-prey interaction, for example, it is good to quickly evolve an extra burst of speed.[5] Biologists Bergstrom and Lachmann (2003) were the first to describe what they call the *Red King effect* using evolutionary game theoretic models. As they show, in species engaged in mutualistic interactions, counter-intuitively, slow evolving species can sometimes gain an advantage.

One feature of many mutualisms is that there are a variety of ways to 'split' or share out the benefits of the mutualism. Consider ants who farm aphids, for example. The aphids reap benefits related to protection from predators, and food supply. In return, they secrete a sugary nectar for the ants. But how much nectar does each aphid provide for how much protection? As Bergstrom and Lachmann (2003) show, sometimes the species that evolves slower does better. This occurs when a fast evolving species evolves accommodating behaviors, which their mutualistic partners then come to take advantage of. One intuitive way to understand this effect is via appeal to an analogous rational choice situation described by Schelling (1980). Suppose two people are playing chicken with their cars. They speed towards each other. Neither wishes to be the one to swerve, but if they both keep going straight, they crash. A way to win this game is to visibly throw your steering wheel out the window. Any sane opponent will then decide to swerve, forfeiting the game. In this case, an individual makes themselves incapable of changing strategies, and gains an advantage. We can think of the Red King effect as resulting from a kind of intransigence that forces a mutualistic partner to swerve in evolutionary time.

How does this relate to minority disadvantage? Bruner (2017) uses evolutionary game theory to derive a formally similar effect in a cultural evolutionary scenario where two groups develop bargaining norms, and one group is in the minority. This occurs as a result of a fundamental asymmetry between the interactive experiences of individuals in minority and majority groups. Minority members, by dint of

[5] This hypothesis is named after the Red Queen in Lewis Carrol's *Through the Looking Glass*. She tells Alice, "Now *here*, you see, it takes all the running you can do to keep in the same place".

small size, are constantly meeting their out-group for interaction. The majority, on the other hand, only rarely meets their out-group. This means that we should expect minority members to much more quickly learn how to interact with their out-group than vice versa. In other words, when it comes to cultural evolution, minority groups will be fast-evolving ones. As Bruner (2017) shows, this can lead to minority disadvantage. A number of papers have tested the robustness of this effect across different games, interactive structures, learning strategies, and modeling choices (Bruner 2017; O'Connor and Bruner 2017; O'Connor 2017; Rubin and O'Connor 2018; O'Connor et al. 2017). As these authors show, the *cultural Red King effect* emerges robustly in evolutionary models as a direct result of the minority-majority asymmetry just described. It is this body of literature that inspires the experimental work presented here.

It will be useful, for what follows, to develop a deeper understanding of this effect using evolutionary game theory as a framework. First we can ask: what sorts of strategic situations are the ones where the cultural Red King effect potentially occurs? To see the cultural Red King effect, we need a strategic scenario where 1) two groups culturally interact and develop conventions 2) there are multiple evolutionary outcomes or *equilibria* that can emerge and 3) there is a conflict in preferences over these equilibria for the two groups involved.

Figure 1 shows an example of a game that meets these conditions. It is often called the *Nash demand game*, and is the primary game used by all the authors who have modeled the cultural Red King to date.[6] The idea is that two individuals bargain to divide a resource. Each demands some portion of it. If they make compatible demands, they each get what they asked for. If their demands are incompatible in that they over-demand the resource, it is assumed that they are unable to peaceably split the resource, and each gets a low payoff.

In particular, this figure shows a 'mini-game' where we assume there is a resource of 10, and each actor can only make a Low, Med, or High demand for 4, 5, or 6 of the resource.[7] This minimal structure is a simple way to capture a scenario where there are multiple ways to divide a resource which may favor either of the

[6] This game was first introduced by Nash (1950), and has subsequently been used to represent countless scenarios of resource division—salary decisions, international trade agreements, bargaining over household labor, etc. It has also been called divide the dollar, divide the pie, divide the cake, the Nash bargaining game, or just the bargaining game.

[7] Sigmund et al. (2001) introduce the term 'mini-game' and defend the usefulness of small, simple game theoretic models. In the case of evolutionary bargaining models there is a long history of using mini-games to capture bargaining scenarios as in Skyrms (1994); Axtell et al. (2000).

		Player 2		
		Low	**Med**	**High**
Player 1	**Low**	4,4	4,5	4,6
	Med	5,4	5,5	0,0
	High	6,4	0,0	0,0

FIGURE 1. A payoff table for a Nash demand game with demands 4, 5, and 6 and a resource of 10. Payoffs for player 1 are listed first.

players involved. Rows represent the strategies for player 1 and columns for player 2. Each entry in the table represents the payoffs for one outcome with player 1 listed first. So, for example, if player 1 demands 6 and player 2 demands 4, they get 6 and 4 respectively. If they both demand 6, they over-demand the resource and each receives a low payoff of 0.

There are three pure strategy Nash equilibria of this game—the outcomes where player 1 demands High and player 2 Low, where player 1 demands Low and 2 High, and where both players demand Med.[8] At these outcomes, the entire resource is perfectly divided. If either player demands more they exceed the resource and get nothing. If either demands less, they get less. Notice that one of these equilibria, the Med-Med one, is 'fair' in the sense that each player gets the same payoff. The other two are 'unfair' in that one player gets more. If we imagine two groups who learn a convention where one side always demands High and the other Low, we have a simple representation of something like discrimination—individuals treat members of the two groups differently, to the detriment of one group.

Let us translate this game into an evolutionary model. Suppose we have two groups that make up one population. Let the majority group constitute proportion p of the population and the minority group $1 - p$ where $p \geq .5$. For each individual, suppose that they condition their strategy on the group membership of the individual they interact with. For instance, Karen might always play Low with in-group members and High with out-group members. The work cited above has shown that in most evolutionary models such a population will evolve to one of three stable outcomes between the two groups, which reflect the Nash equilibria of the underlying game. We can think of these as representing culturally evolved conventions—patterns of behavior that emerge over time to solve social problems.

[8] Pure strategies are those where the actors always take the same action rather than probabilistically mixing. There is also a mixed strategy equilibrium of this game, but it does not emerge in the evolutionary models we will describe.

In the 'fair' convention, everybody demands Med of their out-group. In the two discriminatory conventions, one side always demands High when they meet an out-group member, and the other side always demands Low.[9]

We can now say much more precisely what the cultural Red King effect looks like in an evolutionary game theoretic model. As a group's size $(1-p)$ gets smaller, if their likelihood of ending up at a lower payoff equilibrium increases, we see the cultural Red King. (In a minute, it will become clear what we mean by 'likelihood' here.) Again, this occurs because in bargaining scenarios, low, accommodating demands are broadly successful. Regardless of what your opponent is doing, you guarantee yourself some payoff by being accommodating. This means that in many cases, both groups will start to learn to accommodate. But since minority groups learn this lesson more quickly, at some point the majority group begins to learn that they can do better by taking advantage of this accommodating behavior. Eventually this process tends to lead to conventions of bargaining that advantage the majority.

Figure 2 shows this effect for a particular model. We run simulations of two groups culturally evolving to play the Nash demand game from figure 1. We employ the most commonly used dynamics in evolutionary game theory, the replicator dynamics, to model change.[10] An understanding of the details of these dynamics is not crucial here, but they work by expanding strategies that do well and contracting those that do poorly.[11] Each data point in this figure shows, for some value of p, how often each possible equilibrium emerged. As is evident, as p increases three things happen. First, the fair equilibrium becomes less likely to emerge. Second, the equilibrium where the minority population discriminates becomes less likely. And third, the equilibrium where the majority discriminates becomes increasingly likely. It should be clear now, what it means for a group

[9] Notice that these three outcomes are the population level equilibria between groups in the model described. I.e., in a population playing these strategies, no individuals may change their between-group strategy and improve their payoff. In these models, conventional behavior also emerges within each of the two in-groups. Usually this involves everyone treating each other fairly. Sometimes 'fractious' patterns emerge where some individuals demand High and others Low (Skyrms 1994).

[10] Taylor and Jonker (1978) introduce the replicator dynamics to model evolution by natural selection, but they have also often been used as a good representation of cultural change. In particular, they are equivalent to the average expected population change of explicitly cultural evolutionary and learning models (Weibull 1997; Börgers and Sarin 1997; Hopkins 2002). Since this is a simulation, we employ the discrete time version of the dynamics adapted to a two population model where all individuals interact and the populations are of different sizes.

[11] For more on this particular model see O'Connor (2018a).

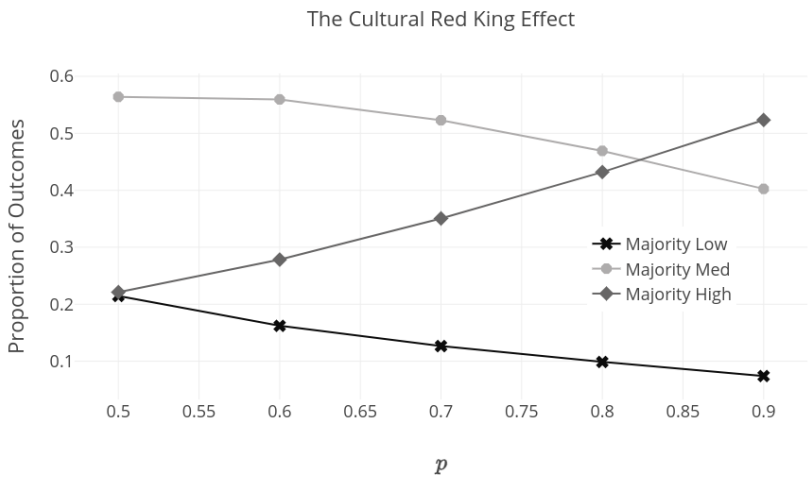


FIGURE 2. A plot showing the cultural Red King effect for a Nash demand game with demands 4, 5, and 6. The x-axis tracks majority proportion. Each line tracks the proportion of simulations that ended up at each of three possible equilibria. As majority size increases, the minority is increasingly disadvantaged.

to become more likely to end up disadvantaged as a result of the cultural Red King—it is a more probable outcome of an evolutionary scenario.

This effect is of obvious interest from a cultural standpoint. The models indicate that minority groups may be prone to disadvantage in situations of resource division and bargaining by dint of size alone. Of course, real world discrimination involves factors that go far beyond what is represented in these models including psychological phenomena related to bias, stereotyping, and stereotype threat (Ogbu 1978; Stewart 2010). This said, the cultural Red King effect may nonetheless contribute to discrimination. The modeling work discussed is also important in showing how little is needed to generate discrimination. For instance, anti-bias training may not be enough to ensure equity when basic cultural evolutionary processes, like the cultural Red King, lead to inequity (Stewart 2010; O'Connor 2018a,b).

There is another result to mention, which is that for different versions of this model we can actually see minority advantage. Some payoff structures mean that it makes the most sense for minority members to quickly learn high demands. Once they do so, majority members do best to make complementary low demands. This *cultural Red Queen* is described at further length by Bruner (2017). However, O'Connor (2017, 2018a) argues that under common conditions, including risk aversion, out-group bias, and pre-existing discriminatory norms, the culture Red

		Player 2	
		Low	**High**
Player 1	**Low**	4,4	4,6
	High	6,4	0,0

FIGURE 3. The payoff table used in the experiment. Payoffs are listed with player 1 first.

King effect is stronger. This increases worries that the cultural Red King might influence real world populations.

3.2. **Predictions.** Let us now turn to the particular model we employ in this experiment and the relevant theoretical predictions. Our goal was to create the most highly simplified set-up that could still reproduce the cultural Red King. In particular, we wanted to isolate the relevant effect by reducing the chances that cultural norms and cognitive biases influenced participant behavior.

One choice along these lines involved using a Nash demand game where the 'fair' demand was removed. Humans show strong inclinations towards fair behavior in laboratory experiments, even when real-world behavior is far from fair (Fehr and Schmidt 1999). There is some debate over whether this inclination derives from cultural norms for fairness, or an innate aversion towards inequity (Binmore and Shaked 2010; Fehr and Schmidt 2010). Whatever the root of this behavior, we force participants to choose between unequal demands to avoid the psychological pull of fairness. The payoff table we use is displayed in figure 3. The particular demands—4 and 6—were chosen as ones where the cultural Red King effect should be particularly strong.[12] There are two pure Nash equilibria of this simplified game. These are the strategy pairings where player 1 demands 6 and player 2 demands 4 and vice versa. In evolutionary models, we thus expect groups to evolve to one of the discriminatory outcomes described where one group always demands 6 and the other 4. Of course, excluding the fair demand from the game here means that the results may not apply as directly to real world scenarios where fairness is an option.

We also chose an interactive structure that was as simple as possible. Above we have described models where individuals interact with both in-group and out-group members and condition their strategies based on the group membership of their partners. Here we only consider between-group interactions. In other words, subjects only interact with their out-group and never their in-group. (As

[12] See Bruner (2017); O'Connor (2017) for a more in-depth explanation.

will become clear in the next section, this meant that majority group members would not interact in every round.) This choice creates a scenario that replicates asymmetries in interaction and learning speed between groups, while avoiding the complications involved in learning both in- and out-group strategies.

The body of literature cited in 3.1 predicts a cultural Red King effect for a majority and minority group learning to interact in many bargaining games. This prediction is based on a large class of models, employing various learning dynamics and population structures. This said, we can also verify that the cultural Red King effect arises in models tuned to the particular parameters used in this experiment. To that effect, we have produced an agent-based model where agents play the game from figure 3. We represent eight agents (again matching the experimental protocol) and vary the size of the minority group.[13] Eventually this model always evolves to one of the equilibrium states—the majority group always demands 6 and the minority 4, or vice versa. And, indeed, the larger the majority group, the greater the chance they evolve to demand 6. The results of this simulation are shown in 4. As described in the next section, we ran experiments of groups with 6 majority members and 2 minority members. This majority group size results in 65% of simulations evolving to the majority advantage equilibrium for this model.

We emphasize, though, that one should not take the particular value here very seriously. The match between the model and real human learning behavior is not a tight one. Rather, this little simulation is simply intended to demonstrate to the reader that the cultural Red King effect is indeed expected to emerge in a model with parameter settings that are similar to those we employ in the experiment. As mentioned, the prediction for this experiment is drawn from a broad set of results that are robust over many models making different assumptions.

Notice that this model, and previous ones, predict that each run ends up at a very clear convention where all members of a group take the same strategy. We do not expect such homogeneity in real groups. Human strategic behavior is highly stochastic. Instead, previous experiments using the laboratory to test

[13] In this particular model, we give each agent a limited memory of past interaction. Each round we randomly choose agents for interaction. They select whatever strategy would perform best against their memories. For example, an agent with many memories of their out-group playing High will do best to respond by playing Low. This agent-based model is based off an influential analytic modeling paradigm used by Young (1993) to investigate bargaining between groups. Axtell et al. (2000) develop an agent-based version of the model, and O'Connor (2017) demonstrates the cultural Red King effect in this agent-based version. The choice of dynamics here is somewhat arbitrary. We used this learning rule since it has been previously used to explore the cultural Red King, but results will be qualitatively robust across many rules.

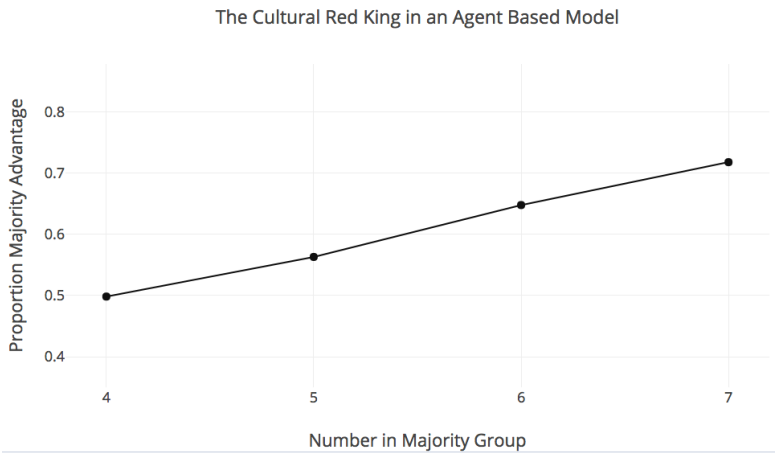


FIGURE 4. Results from a simulation of an agent-based model demonstrating the cultural Red King effect for parameter values fit to the experimental settings. The results show the proportion of runs of simulation that end up at a convention where the majority group demands 6 and the minority 4.

evolutionary game theoretic predictions have yielded results that are messy but generally match trends in evolutionary models, rather than perfect model-world correlation.[14]

We can now state our main theoretical prediction.

Prediction 1: We predict that for experimental subjects playing the game in figure 3 and sorted into a minority and majority group, the majority individuals will demand High more often, and receive higher payoffs than the minority individuals as a result of the cultural Red King effect.

We make one further prediction. The cultural Red King emerges, in all models studied, over the course of learning. Therefore we should expect in our experiments for the difference in frequency of demanding Low between the two groups to increase as the experiment continues.

Prediction 2: We predict that the difference in frequency of demands between the two groups will increase over the course of the experiment, with the minority receiving relatively lower payoffs at the end than at the beginning.

[14] See, for example, Blume et al. (1998); Bruner et al. (2018) who find that subjects in the lab conform to evolutionary predictions in the signaling game, but by no means accord to the perfect conventions predicted.

4. Experimental Setup

Our experiment was programmed in oTree (Chen et al. 2016) and conducted both online using Amazon Mechanical Turk and at the Experimental Social Science Laboratory at the University of California, at Irvine, where participants were drawn from a pool of undergraduate and graduate students.

For the experiment, participants were asked to play the simplified Nash demand game described in section 3.2. The experiment was run for 14 sessions, each of which involved eight participants interacting over 100 rounds of play. In total, 112 individuals participated in the experiment.

The first ten sessions were run online at Amazon Turk. Participants were screened for the experiments by running a preliminary HIT (human intelligence task) which was available only to Amazon Turk workers who had previously completed at least 100 HITs with a 90% approval rate (so as to ensure participant reliability) and which provided a tutorial explaining the structure of the game, and then asked comprehension questions to test whether participants clearly understood the payoffs to themselves and their opponents for each combination of demands. Participants who answered all questions correctly were then invited to participate in the experimental session.

The last four sessions were run in the laboratory. (As we will describe, this was necessitated by a software change on the part of Amazon Turk.) Before each laboratory session, participants were asked to sit at a randomly assigned computer terminal where they would play games with the other participants. Before the session, participants were provided with a tutorial explaining the structure of the game. Just as with the Amazon Turk experiments, individuals were asked comprehension questions to confirm that they had clearly understood the payoffs to themselves and their opponents for each combination of demands. In the lab, we were not able to filter for only individuals who answered all the comprehension questions correctly. That said, the risk for inattention in Amazon Turk participants is typically far greater than those in the laboratory, and most participants (91%) in the laboratory sessions did in fact answer all questions correctly.

All else was identical in the implementation of the Amazon Turk and laboratory versions of the experiments, with only the difference that participants in the former setup were (typically) at computers in their homes while participants in the latter setup were at computer terminals in the laboratory. Importantly, as we will discuss

in section 5, there was no significant difference in the results of the laboratory and online setups.

At the start of each session, participants were randomly and permanently assigned to either a majority group of six individuals or to a minority group of two individuals. In each round, members of each group were randomly matched with members of the opposing group. There were no in-group interactions. Thus, participants in the minority group played for all 100 rounds while participants in the majority group played an average of 33 rounds. In rounds where majority members did not play, they were given a screen instructing them to wait.

Importantly, individuals were not informed of the group to which they were assigned, or even of the presence of any groups at all. This was done to isolate the desired causal factors. It is well known that framing interactions in terms of groups can elicit strong behavioral responses (Tajfel 1970). We did not wish to invite such responses. Instead, our aim was to isolate the effect of differentials in the frequency of interaction between groups of different sizes on the propensity to arrive at inferior outcomes.

In each round of play, each individual involved in the round chose her action and observed the payoff she received given the combination of her action and the action of her opponent. Having observed her opponent's action as well as both her payoff and her opponent's payoff, each participant could choose to adjust her strategy in the next round of play. Note that, in keeping with the context-free nature of much of experimental economics, we did not refer to these actions as 'demands' during the experiment. Instead, subjects were told they could either 'choose 6' or 'choose 4'. This was done so as to minimize the effect of social norms or contextual rules (e.g. 'avoid being too demanding in your social interactions'), which could influence their decision making.[15] Screenshots of the experiment introduction, instructions, understanding questions, and decision pages can be found in Appendix C.

All participants received a baseline of \$7 for participation in the experiment. Additionally, to incentivize strategic behavior, participants were able to earn more based on their performance. Participants were made aware of this fact and of the details of how this would be done. For each participant, at the conclusion of her 100 rounds of play, five rounds would be selected at random and her average payoff for those five rounds—which could range from \$0 to \$6—would be added

[15] Though, see section 6 for a discussion of how subjects' decision making might still have been affected by inequity aversion.

to her final payment.[16] Participants typically completed all 100 rounds within 30 minutes in both laboratory and online experiments. Online, participants were paid through Amazon within a day of completing the session, and in the laboratory participants were paid in cash immediately after each session.

As mentioned, 10 session were conducted on Amazon Turk and 4 sessions in the laboratory. This was necessary because Amazon Turk changed its software partway through our experiments in a way that made our existing experimental design—one that required simultaneous coordinated interactions between groups of individuals—impractical to implement. There are some demographic differences between the Amazon Turk subject pool and the laboratory subject pool. In short, the Amazon Turk subject pool tends to be more diverse than laboratory subject pools composed of graduate and undergraduate students. Previous studies, though, have shown that Turk studies tend to produce quality data on par with that of the student pools at university laboratories (Buhrmester et al. 2011; Paolacci and Chandler 2014; Kees et al. 2017). Thus, when experimental hypotheses do not depend on the demographic differences that exist between these two pools, studies can be validly performed on one or both subject pools.

Importantly, for our experimental predictions we require only that individuals be responsive to incentives in such a way that they try to improve their payoffs and that they learn through repeated interactions. These requirement are not sensitive to demographic changes between the Amazon Turk and laboratory subject pools and should hold just as well in both cases.

5. Results and Analysis

We will proceed by addressing the theoretical predictions made in section 3. Our first prediction was that groups size differences on their own would drive minority group disadvantage. To test this, we formulate appropriate null and alternative hypotheses as follows:

Prediction 1: Minority Group Disadvantage

H_0 : Minority groups *will not* end up playing 'demand low' with greater mean frequency than majority groups.

[16] Random selection of rounds for payment after an experiment is used to prevent known distortions in subject behavior that occur as participants perceive their level of wealth to change (Davis and Holt 1993).

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\bar{x}_{min}$	0.68	0.9	0.53	0.65	0.48	0.50	0.85	0.8	1	0.78	0.73	1	0.5	0.3
$\bar{x}_{maj}$	0.33	0.44	0.65	0.58	0.44	0.53	0.42	0.47	0.26	0.82	0.46	0.08	0.53	0.71
$\bar{x}_{d}$	0.34	0.46	-0.12	0.07	0.03	-0.03	0.43	0.33	0.74	-0.05	0.27	0.92	-0.03	-0.41

TABLE 1. Frequency of 'demand low' strategy in the final 20 rounds of play (of 100 total rounds) by group over the 14 sessions. The first row displays the frequency for the minority group, the second for the majority group, and the third displays the difference in frequency between the groups. Sessions 1-10 were run on Amazon Mechanical Turk and sessions 11-14 in the ESSL lab at UC Irvine.

H_1 : Minority groups *will* end up playing 'demand low' with greater mean frequency than majority groups.

To assess these, we compare the mean frequency of the 'demand low' strategy for each group in the final twenty rounds of play to test if there is a statistically significant difference.

In nine of our fourteen sessions the minority group ended up playing 'demand low' more frequently than the majority group. The mean frequencies in question are reported for all fourteen sessions in Table 1. (Sessions 1-10 were performed on Amazon Mechanical Turk, and 11-14 in the lab.) In seven of these sessions, the difference in frequency of low demands between the minority and majority group was larger than 0.25. For reference, only once did the majority group end up making low demands with such a differential.[17]

Results are stochastic, but this is what we expect given previous work on game theoretic behavior in the lab. In addition, we expect a variety of outcomes based on our evolutionary models—where even under the cultural Red King the majority group sometimes ends up disadvantaged. The theoretical prediction is that despite this stochasticity we will see a tendency towards disadvantage for the minority group.

Across the sessions, the minority group ended up playing 'demand low' in the final twenty rounds of the sessions with a mean frequency of 0.69 while the majority group played 'demand low' with a mean frequency of 0.50. The mean difference between the minority and majority groups in their frequency of playing 'demand

[17] All of the data from our experiments is registered at the Open Science Framework (`www.osf.io`). The URL to the data repository is located in Appendix A.

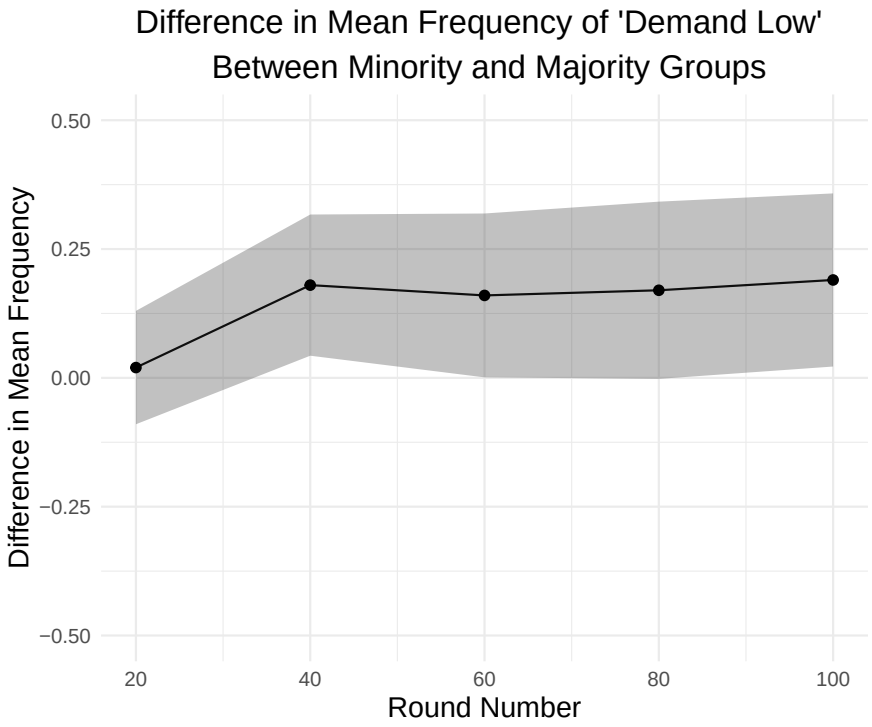


FIGURE 5. Difference in the mean frequency of play of the 'demand low' strategy between the minority and majority groups for each 20-round period of the game across all experiments. Positive values indicate that the minority group played 'demand low' with greater frequency. The gray band displays the 95% confidence intervals.

low' was 0.19 with a variance of 0.14. For our sample of 14 sessions, this yields a p-value of 0.04, a power of 0.58, and a Bayes factor of 3.39 corresponding to conventionally 'substantial' evidential support.

For our statistical test, we employ Student's t-test on the differences of means of the two groups with a standard $\alpha = 0.05$ significance threshold. Thus, for our test, if the null hypothesis is true, the probability of rejecting the null is 5%. And if the alternative hypothesis is true, the probability of failing to reject the null is 42%, given our effect size of 0.19. That is, the data support our theoretical predictions and our result is significant but, given that it is also under-powered, it should be taken as suggestive rather than conclusive.

In figure 5 we see the difference in mean frequency across all sessions of the 'demand low' strategy between the minority and majority groups for each 20-round period of the game. Plots of the evolution of group and individual frequency of play over the course of each session can be found in Appendix B.

A further prediction of our theory was that the difference in mean frequency of playing 'demand low' should increase over repeated interactions. We can make this precise by testing whether the mean difference in the last 20 rounds of play was greater than the first 20 rounds of play.

Prediction 2: Progressive Disadvantage

H_0 : The difference in mean frequency of minority groups and majority groups playing 'demand low' *will not* increase over the course of play.

H_1 : The difference in mean frequency of minority groups and majority groups playing 'demand low' *will* increase over the course of play.

The mean difference between the groups in the first 20 rounds of play was 0.02. That is, both minority and majority groups played 'demand low' in the first 20 rounds with essentially the same frequency. The mean difference between the groups in the last 20 rounds of play was 0.19. That is, the minority groups played 'demand low' more frequently. This amounts to a mean increase of 0.17 with a variance of 0.09 which yields a p-value of 0.027, a power of 0.63, and a Bayes factor of 4.11 corresponding to conventionally 'substantial' evidential support. That is, the difference in play increases over the course of play and the increase is significant.

Addressing our concerns that the online and laboratory populations may be different, we verify that their observed mean behavior is indeed quite close and the difference is not statistically significant. In particular, the mean difference for the online sessions (1-10) and laboratory sessions (11-14) is 0.1 with a standard error of 0.23 yielding a p-value of 0.67. Thus, the evidence does not support a hypothesis that the behavior of individuals in the online and laboratory contexts is substantively distinct.

A close examination of individual play reinforces the insight that there is high variance in individual behavior. Some participants began sessions by playing 'demand low' and persisted in doing so without ever experimenting with a more aggressive strategy even when it would appear to have been beneficial to do so, some began by playing 'demand high' and persisted in doing so even at cost to themselves. Most participants exhibited a slight aversion to aggressive demands as is demonstrated by the greater mean and median frequencies of the 'demand low' strategy. There are multiple plausible, distinct, compatible explanations for these behaviors: picking a single strategy is appealingly easy; some individuals count

on being able to bully others into submission; some individuals are mortified by the thought of disadvantaging anyone; most individuals are disposed, by varying degrees, to abide by (and enforce) equitable norms (Henrich et al. 2001).

This is all true. But what may be most impressive is that, in spite of the endogenous variation in individual behavior, our results show that minority groups do end up at payoff-inferior outcomes with significantly greater frequency. There is a signal in the noise.

6. Conclusion

While we have found support for the cultural Red King hypothesis, there are a few factors to consider which may have influenced our results. In particular, one curious observation is that even when most of the minority group demands low, not all members of the majority group consistently demand high (see, e.g., the graphs for sessions 2, 7, 8, and 9 in Appendix B). The fact that majority group members did not consistently learn to take full advantage of the lower demands made by the minority group may partially explain why the Red King effect we observed was not stronger.

There are a few possible explanations of this observation. While we kept the experiment as context-free as possible, subjects could still have seen that their 'choosing 6' would, when the person they interacted with chose 4, lead to their getting more than the other person. Therefore, subject decision making could still have been affected by inequity aversion or other social norms regarding getting more than one is due, leading majority group members to avoid making high demands. Another possible explanation is that subjects in these majority groups exhibited risk aversion, so that majority group members chose the safer option of demanding low even when they could have maximized payoffs by demanding high.

With respect to the applicability of the results presented here, there are many causes of economic inequity. As mentioned, these include psychological factors like racial and gender biases, and, potentially, stereotype threat. This paper has investigated an additional factor—the cultural Red King—which could potentially lead to minority disadvantage solely due to group size. While this effect cannot be taken to explain any particular case of real world inequity on its own, it is important to recognize that it may, indeed, contribute to inequitable patterns of resource division.

These results are relevant to potential interventions on inequity. For example, many organizations train members in attempts to lessen the effects of racial and gender bias. One worry is that such attempts may not be enough if an effect like the cultural Red King, which results simply from self-interested learning in bargaining scenarios, is at play. In such cases, because the dynamics of interaction tend towards inequity, continued efforts to counteract these dynamics might be necessary.

As mentioned in the introduction, the results here have implications for political and social philosophy. As we pointed out, Nozick's influential account of historical justice argues that distributions of goods that arise from just processes are themselves just. In particular, he argues that such processes must involve a just initial acquisition of holdings, and just transfers, but says relatively little about what this involves. Kenneth Arrow, as early as 1978, introduced the following worry about this account,

> Suppose a dominant group, say whites or"Aryans", agreed to trade with the complementary minority only on very unfavorable terms. Indeed, they might not have to agree in any concrete sense: suppose each one happened for his own reasons to resolve to so act...Are we to say that the results are just? (Arrow 1978, 272).

The worry raised by Arrow is that under Nozick's account, it could be perfectly just to have economic inequity tracking along social identities like race and gender. The modeling work described in this paper, and the experiment presented, deepen this worry significantly. Notice that we can interpret the models here as respresenting joint action—actors produce a good and then divide it. Under this interpretation, they obtain their good in a just way (through their own production) and divide it justly (through a mutually agreeable bargain).

The problem then is not just that one group could *possibly* agree to only trade with another unfavorably. Rather, under completely ubiquitous conditions this sort of inequitable bargaining emerges naturally, and with high probability. The dynamics of human interaction push us towards just this sort of inequity. It is not that this could happen because of some bad actors, but that it likely will happen whenever people recognize social categories, and learn to bargain together, even absent racial or gender bias. Furthermore, features of the social world like minority group size turn out to be highly relevant to determining ultimate distributions of goods.

We do not think this worry will necessarily sway those highly committed to Nozick's libertarian principles.[18] Rather, we point out that one cannot hold to historical justice, while also holding that gender and race should not be important determinants of economic distributions. Historical processes tend to make them so, unless we intervene. Philosophers will have to choose.

There is another more general lesson here for political philosophers who think that establishing just processes is enough to guarantee justice. The lesson is that it is important to actually investigate how social processes proceed, either through models or experiments or both. It is counter-intuitive that group size alone could influence distributions significantly, but the research presented here indicates that it can. In other words, it may not be possible to fully understand the implications of various processes from the armchair. One take-away is that if the outcomes resulting from seemingly just processes tend to be inequitable in surprising ways, it might not be possible to label processes as just without considering their actual consequences.

Acknowledgements

This research was supported by NSF grant 1535139. Thanks to the UC Irvine Social Dynamics seminar for feedback, to the ESSL lab at UC Irvine, to Nicholas Smith, Mike Ashfield, and to Michael McBride. Many thanks to participants at the California Philosophy Workshop for comments, especially Wendy Salkin, Joshua Armstrong, David Plunkett, Gabbrielle Johnson, and Karl Schafer. Special thanks to Justin Bruner for inspiration, and for feedback.

[18] This said, there are a few points that strengthen the worry. First, Nozick adheres to a 'Lockean proviso' that individuals cannot leave interactive partners worse off than they started. In these models, interpreted as involving joint action, individuals always benefit from interaction. It is just that some may benefit more than others. Second, one might worry that equitable distributions are unstable, because the very talented will be able to (fairly) garner high payments for their services. This may generally be true, but in the world of the models we consider here fairness is a stable possibility. Unfairnesses emerges nonetheless.

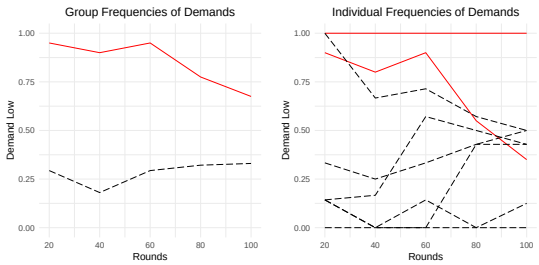
Appendix A. Experiment Data

All of the data from our experiments is registered at the Open Science Framework and can be found, anonymized for review, at the following URL: `https://osf.io/mtc9f/`

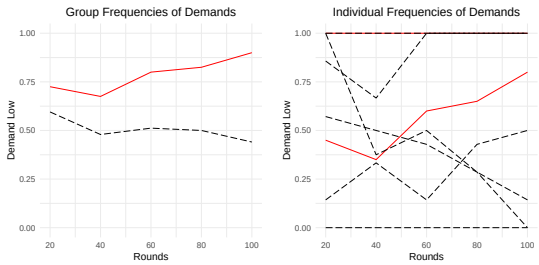
Appendix B. Plots of Individual and Group Behavior

Legend: Minority Group (—); Majority Group (- - -).

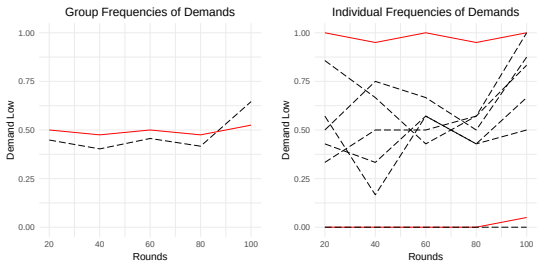
Session 01



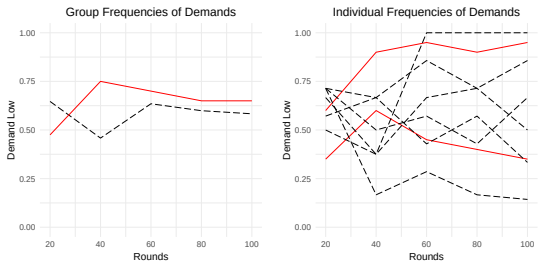
Session 02



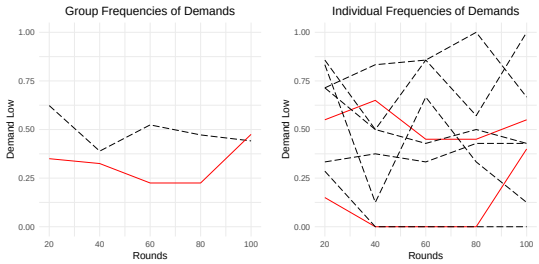
Session 03



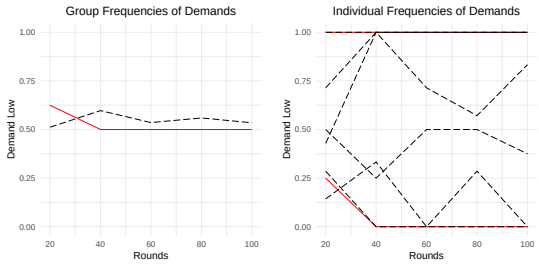
Session 04



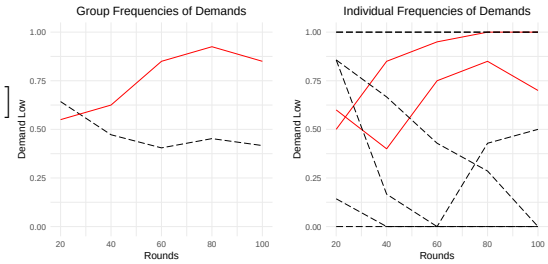
Session 05



Session 06

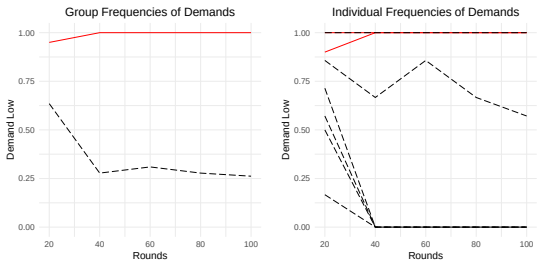


Session 07

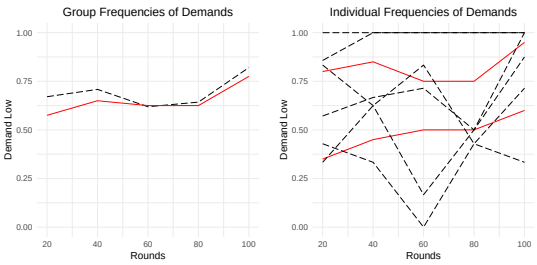


Session 08

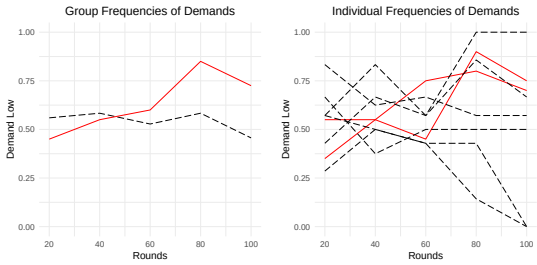
Session 09



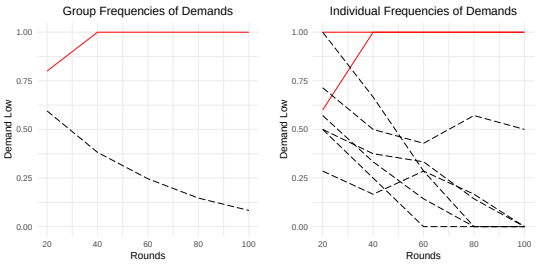
Session 10



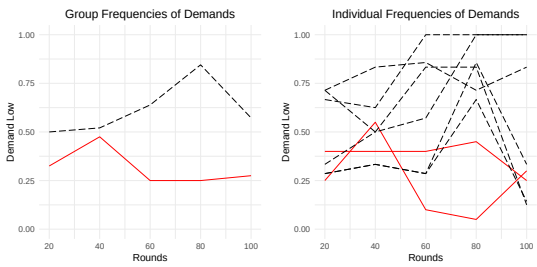
Session 11



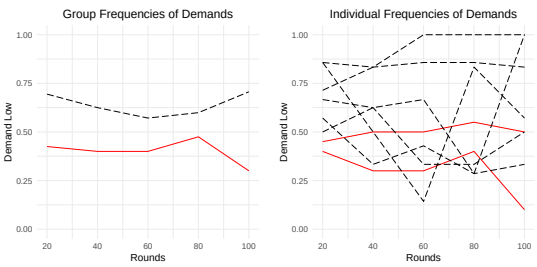
Session 12



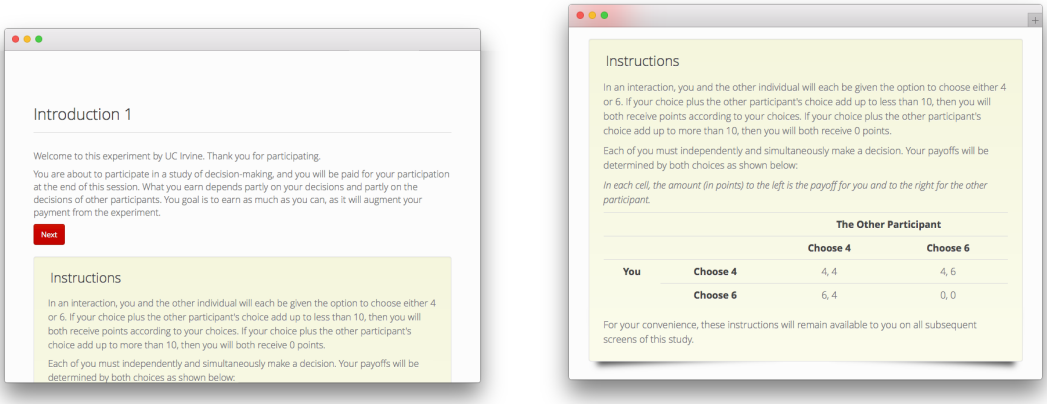
Session 13



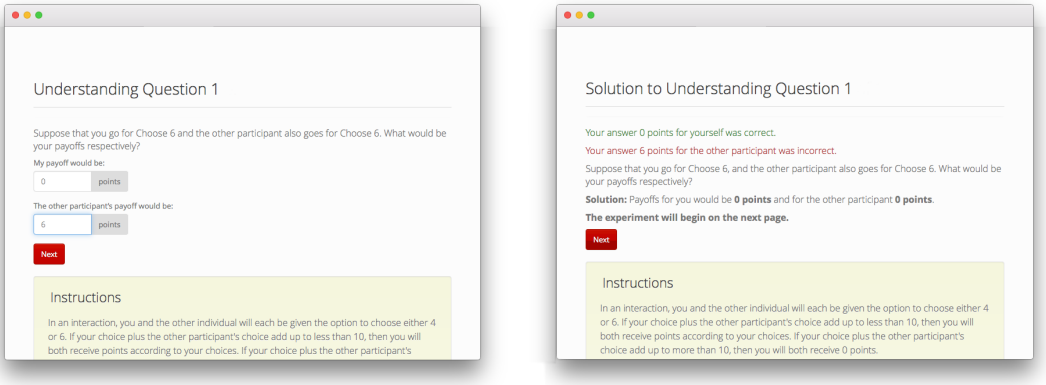
Session 14



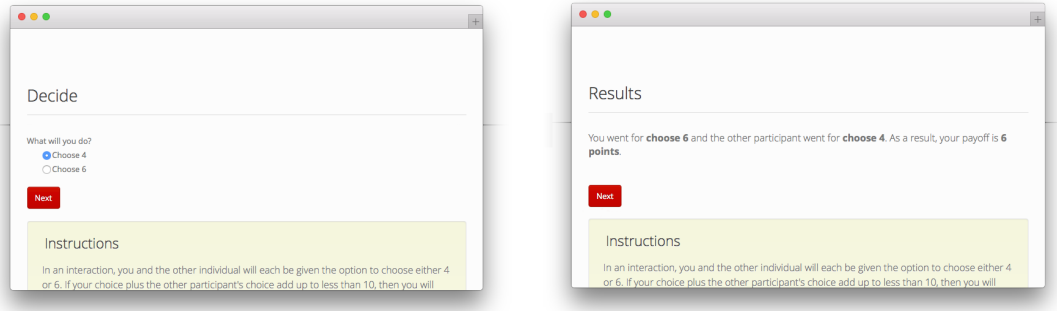
Appendix C. Screenshots of the Experiment Interface



The first introduction and instructions screens.



The first understanding question and question solution screens.



The in-game decision and decision results screens.

References

Adams, J. S. and S. Freedman (1976). Equity theory revisited: Comments and annotated bibliography. *Advances in experimental social psychology 9*, 43–90.

Arrow, K. J. (1978). Nozick's entitlement theory of justice. *Philosophia 7*(2), 265–279.

Axtell, R., J. M. Epstein, and H. P. Young (2000). The emergence of classes in a multi-agent bargaining model.

Bergstrom, C. T. and M. Lachmann (2003). The red king effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences 100*(2), 593–598.

Binmore, K. and A. Shaked (2010). Experimental economics: Where next? rejoinder. *Journal of economic behavior & organization 73*(1), 120–121.

Blume, A., D. V. DeJong, Y.-G. Kim, and G. B. Sprinkle (1998). Experimental evidence on the evolution of meaning of messages in sender-receiver games. *The American Economic Review 88*(5), 1323–1340.

Börgers, T. and R. Sarin (1997). Learning through reinforcement and replicator dynamics. *Journal of Economic Theory 77*(1), 1–14.

Bruner, J., C. O'Connor, H. Rubin, and S. M. Huttegger (2018, Feb). David lewis in the lab: experimental results on the emergence of meaning. *Synthese 195*(2), 603–621.

Bruner, J. P. (2017). Minority (dis) advantage in population games. *Synthese*, 1–15.

Buhrmester, M., T. Kwang, and S. D. Gosling (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science 6*(1), 3–5. PMID: 26162106.

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

Cooper, D. J. (2014). A note on deception in economic experiments. *Journal of Wine Economics 9*(2), 111–114.

Croson, R. (2005). The method of experimental economics. *International Negotiation 10*(1), 131–148.

Davis, D. and C. Holt (1993). Experimental Economics: Methods, Problems, and Promise. *Estudios Económicos 8*, 179–212.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics 114*(3), 817–868.

Fehr, E. and K. M. Schmidt (2010). On inequity aversion: A reply to binmore and shaked. *Journal of economic behavior & organization 73*(1), 101–108.

Freedman, S. M. and J. R. Montanari (1980). An integrative model of managerial reward allocation. *Academy of Management Review 5*(3), 381–390.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath (2001). In search of Homo economicus: Behavioral experiments in 15 small-scale societies. *American Economic Review 91*(2), 73–84.

Hopkins, E. (2002). Two competing models of how people learn in games. *Econometrica 70*(6), 2141–2166.

Kees, J., C. Berry, S. Burton, and K. Sheehan (2017). An analysis of data quality: Professional panels, student subject pools, and amazon's mechanical turk. *Journal of Advertising 46*(1), 141–155.

Nash, J. (1950). The bargaining problem. *Econometrica: Journal of the Econometric Society*, 155–162.

Nozick, R. (1974). Anarchy, state, and utopia. *New York basic book*.

O'Connor, C., L. K. Bright, and J. P. Bruner (2017). The emergence of intersectional disadvantage. Working Paper.

Ogbu, J. U. (1978). Minority education and caste: The american system in cross-cultural perspective.

O'Connor, C. (2017). The cultural red king effect. *The Journal of Mathematical Sociology 41*(3), 155–171.

O'Connor, C. (2018a). *The evolution of inequity*. Oxford, England: Oxford University Press. forthcoming.

O'Connor, C. (2018b). Modeling minimal conditions for inequity. pp. unpublished.

O'Connor, C. and J. Bruner (2017). Dynamics and diversity in epistemic communities. *Erkenntnis*, 1–19.

Paolacci, G. and J. Chandler (2014). Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science 23*(3), 184–188.

Rubin, H. and C. O'Connor (2018). Discrimination and collaboration in science. *Philosophy of Science 85*(3), 380–402.

Schelling, T. C. (1980). *The strategy of conflict*. Harvard university press.

Sigmund, K., C. Hauert, and M. A. Nowak (2001). Reward and punishment. *Proceedings of the National Academy of Sciences 98*(19), 10757–10762.

Skyrms, B. (1994). Sex and justice. *The Journal of philosophy*, 305–320.

Smith, V. L. (1976). Experimental economics: Induced value theory. *The American Economic Review 66*(2), 274–279.

Stewart, Q. T. (2010). Big bad racists, subtle prejudice and minority victims: An agent-based analysis of the dynamics of racial inequality. In *Annual Meeting of the Population Association of America*.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American 223*(5), 96–102.

Taylor, P. D. and L. B. Jonker (1978). Evolutionary stable strategies and game dynamics. *Mathematical biosciences 40*(1-2), 145–156.

Tilly, C. et al. (1998). *Durable inequality*. Univ of California Press.

Weibull, J. W. (1997). *Evolutionary game theory*. MIT press.

Wylie, A. (2013). Why standpoint matters. In *Science and other cultures*, pp. 34–56. Routledge.

Young, H. P. (1993). An evolutionary model of bargaining. *Journal of Economic Theory 59*(1), 145–168.

E-mail address: amohseni@uci.edu

E-mail address: cailino@uci.edu

E-mail address: hannahmrubin@gmail.com