Supplemental Appendix A

Institutional Sources of Legitimate Authority: An Experimental Investigation

Sample Instructions for Subjects (Salary/Full Information Treatment)
Instructions

This is an experiment on decision making. In the following experiment you will make a series of choices. At the end of the experiment, you will be paid depending on the specific choices that you made during the experiment and the choices made by other people. If you follow the instructions and make appropriate decisions, you may make an appreciable amount of money. Please remain silent and listen carefully to the instructions.

In addition to the show-up fee of $7, during the course of the experiment you will have the opportunity to earn “tokens” that will be converted into dollars at the end of the experiment. The conversion rate is:

40 tokens = 1 dollar

You will be assigned, at random, into a group of five people. Within that group, you will also be randomly assigned to one of two roles in the experiment: Role A, or Role B. Within a given group, one person is randomly assigned to Role A; the other four people are assigned to Role B. The assignments will remain fixed for the duration of the experiment: that is, you will interact with the same group of other people, and remain in the same role, for the duration of the experiment. In addition, every person assigned to Role B will also receive an ID number: 1, 2, 3, or 4. This ID number, unlike groups and roles, will be randomly re-assigned from one period to the next. All of your interactions will be through the computer terminals at which you are sitting, and your true identity will never be revealed to any other person in the laboratory.

The experiment consists of 20 periods, all of which have the same structure. In each period, there are up to three separate stages. In the first stage of each period, each person in Role B will be given a supply of tokens, and must choose whether to allocate these tokens to a common pot or to keep these tokens for him- or herself. In the second stage of each period, the person in Role A will receive information about the individual choices made in the first stage by each of the people in Role B (listed by their ID number for the period). The person in Role A will then have an opportunity, if he or she wishes to do so, to choose one person in Role B and to attempt to reduce the payoffs of that person. If the person in Role A does attempt to reduce the payoffs of a person in Role B, a third stage takes place. In the third stage, all of the other people in Role B can choose to increase the likelihood with which the person in Role A is successful in reducing this person’s payoffs, to decrease this likelihood, or to leave it unchanged.

This same process will be repeated in all 20 periods. A more complete description of this process now follows.

First Stage

At the beginning of each period, each person in Role B receives 20 tokens. Each person in Role B must then decide whether to allocate these 20 tokens to a common pot, or to
keep them for him- or herself. Each person in Role B must choose either to allocate all 20 of the tokens or keep all 20 of the tokens.

The payoff to a person in Role B in the first stage is composed of two parts:

- The number of tokens the person keeps for him- or herself PLUS
- 0.4 times the number of tokens that all people in Role B allocate to the common pot (including tokens the person allocates him- or herself).

That is, the payoff to a person in Role B in the first stage can be written as

$$\text{first-stage payoff to person in Role B} = (\text{tokens kept}) + 0.4 \times (\text{total tokens allocated to common pot by people in Role B}).$$

Although the person in Role A does not make a choice in the first stage, he or she also receives first-stage payoffs that depend on the choices made by people in Role B. These payoffs are composed of two parts:

- An automatic supply of 20 tokens PLUS
- 0.4 times the number of tokens that all people in Role B allocate to the common pot.

$$\text{first-stage payoff to person in Role A} = 20 + 0.4 \times (\text{total tokens allocated to common pot by people in Role B}).$$

Therefore, every token kept by a given person in Role B increases that person’s first-stage payoffs by one token (and does not contribute to the payoffs of other group members). Every token allocated to the common pot increases the first-stage payoffs of every person in Role B by 0.4 tokens, and also increases the first-stage payoffs of the person in Role A by 0.4 tokens.

Consider the following examples:

- Suppose that every person in Role B keeps all of his or her 20 tokens for him- or herself. Then the first-stage payoffs of each person in Role B will be equal to $20 + (0.4 \times 0) = 20$ tokens. The first-stage payoffs of the person in Role A will be equal to $20 + (0.4 \times 0) = 20$ tokens.
- Suppose that every person in Role B allocates all of his or her 20 tokens to the common pot. Then the first-stage payoffs of each person in Role B will be equal to $0 + (0.4 \times 80) = 32$ tokens. The first stage payoffs of the person in Role A will be equal to $20 + (0.4 \times 80) = 52$ tokens.
- Suppose that two people in Role B each allocate their 20 tokens to the common pot, and two people in Role B each keep their 20 tokens for themselves. In total, 40 tokens are allocated to the common pot by the people in Role B as a whole. Then each person in Role B receives $0.4 \times 40 = 16$ tokens.
tokens from the common pot. The two people in Role B who each allocated 20 tokens to the common pot (keeping none for themselves) would therefore have first-stage payoffs equal to \(0 + (0.4 \times 40) = 16\) tokens. The two people in Role B who each allocated no tokens to the common pot (keeping 20 for themselves) would therefore have first stage payoffs equal to \(20 + (0.4 \times 40) = 36\) tokens. The person in Role A would receive first stage payoffs equal to \(20 + (0.4 \times 40) = 36\) tokens.

Second Stage

At the beginning of the second stage, everyone receives some information about what happened in the first stage. Specifically, both the person in Role A and each person in Role B is told:

- What the decision of each person in Role B was in the first stage, listed by his or her ID number for the period (1, 2, 3, or 4). That is, everyone receives a message about whether each person in Role B “allocated” his or her tokens to the common pot or “kept” his or her tokens.
- His or her first-stage payoffs.

Once this information has been received, the person in Role A must decide whether he or she wishes to choose one person in Role B (say, the person with ID number 2 for the period), and attempt to reduce his or her first-period payoff. An attempt to reduce payoffs in this way will be successful with a likelihood that depends on choices made by the other people in Role B in the third stage. Attempting to reduce the payoffs of a person in Role B costs the person in Role A 2 tokens, whether the attempt is successful or not.

If the person in Role A makes an attempt and it is successful:

- The payoffs of the chosen person in Role B are reduced by 30 tokens. The person in Role A is not able to adjust this amount.
- The person in Role A does not get to keep any tokens he or she deducts from the payoffs of the person in Role B.
- The person in Role A pays the 2 token cost of making the attempt.

If the person in Role A makes an attempt and it is unsuccessful:

- The payoffs of the chosen person in Role B are unchanged.
- The payoffs of the person in Role A are unchanged, except for the 2 token cost of making the attempt.

After the person in Role A has made his or her decisions, each person in Role B will learn whether or not the person in Role A has attempted to reduce the payoffs of any person in Role B, and if so, which person this was.
**Third Stage**

The third stage takes place only if the person in Role A attempted to reduce the payoffs of a person in Role B. In this stage, the people in Role B who were not selected by the person in Role A have the opportunity to change the likelihood that the person in Role A’s attempt to reduce the payoffs of the person in Role B will be successful.

Imagine a jar containing six balls, which can be either red or white. Whether or not an attempt to reduce the payoffs of the selected person in Role B is successful will be determined by the color of the ball drawn from such a jar, at random, by the computer. If a red ball is drawn, the attempt will be successful, and the payoffs of the selected person will be reduced. If a white ball is drawn, the attempt will be unsuccessful, and the payoffs of the selected person will remain unchanged.

Initially, before the other people in Role B make their decisions in this stage, three of the balls are red, while three of the balls are white.

In the third stage, each person in Role B who was not selected by the person in Role A has three choices:

- Pay a cost of 1 token to replace one of the white balls in the jar with a red ball.
- Pay a cost of 1 token to replace one of the red balls in the jar with a white ball.
- Pay no cost, and do not replace a ball.

As such, if the person in Role A attempts to reduce the payoffs of a person in Role B, the final likelihood of success depends on the decisions made by the other people in Role B. For example:

- If all three of the other people in Role B choose to replace a white ball with a red ball, then the final likelihood of success will be 100% (because all of the balls in the jar would be red).
- If all three of the other people in Role B choose to replace a red ball with a white ball, then the final likelihood of success will be 0% (because all of the balls in the jar would be white).
- If all three of the other people in Role B choose not to change the color of any of the balls, then the final likelihood of success will be 50% (because three of the balls in the jar would be red, while the other three would be white).
- If two of the other people in Role B choose to replace a white ball with a red one, while one chooses to replace a red ball with a white one, then the final likelihood of success will be two-thirds, or $66\frac{2}{3}$% (because four of the balls in the jar would be red, while the other two would be white).

At the end of the third stage, the person in Role A and all people in Role B will learn what choices were made by people in Role B in the third stage. They will also learn whether or not the attempt by the person in Role A to reduce the payoffs of a person in Role B was successful.
Summary of Net Payoffs for a Period

For a person in Role B, the following is calculated:

- First-stage payoffs (from common pot and tokens kept) ….
- …MINUS the number of tokens, if any, spent in the third stage
- …MINUS the number of tokens, if any, that were ultimately successfully reduced by the person in Role A

The result is the net payoffs for the period for a person in Role B.

For the person in Role A, the following is calculated:

- First-stage payoffs (from common pot and automatic token supply)
- …MINUS the number of tokens, if any, spent in the second stage

The result is the net payoffs for the period for the person in Role A.

Regardless of your Role, you will see your net payoffs for the period on your screen once that period is complete.

Conclusion

This concludes the description of the choices that are made and the payoffs that are earned in one period. This process will be repeated until all of the 20 identical periods are completed. Your payment for the experiment will consist of the sum of your payoffs from all 20 periods, plus the show-up fee. Remember that you will interact with the same group of other people, and remain in the same role, throughout this process. Also remember that the ID numbers of people in Role B are randomly reassigned from one period to the next.

We ask everyone to remain silent until the end of the last period and then to await further instructions. If you have any questions, please raise your hand while remaining silent.
Supplemental Appendix B.

A Model of Citizen Opportunities to Assist or Hinder Government

1 Model Primitives

There are \( n + 1 \) players: an authority figure \( a \) and \( n \) citizens indexed by \( i = 1, \ldots, n \). The stage game unfolds as follows: citizens begin each stage with an endowment \( y_c \) (common to all citizens) and the authority with endowment \( y_a \). The sequence of moves proceeds as follows:

1. Each citizen makes a binary public goods contribution decision

2. The authority observes contribution choices and decides which citizen, if any, to target for enforcement

3. With probability \( \gamma \), citizens observe each others’ contribution choices

4. Untargeted citizens observe the leader’s choice of whether and whom to target, and decide whether to intervene on behalf of the leader, on behalf of the enforcement target, or neither.

Citizen contributions go to a standard linear public good with marginal rate of return \( r \in (1/n, 1) \). Let \( C_i \) be an indicator equal to one if citizen \( i \) contributed, and zero otherwise. The contribution amount (if made) equals the citizen’s endowment \( y_c \).

The fact that the authority can target only a single citizen is meant to capture the fact that government capacity is limited. Targeting a single citizen costs the authority \( e < y_a \). Let \( E_i \) be an indicator equal to one if the authority attempts to enforce against citizen \( i \), and \( E \) be an indicator equal to one if the authority targets any citizen. Note that we do not assume that a citizen must have failed to contribute in order to be targeted. In other words, “innocent” and “guilty” citizens are both susceptible to enforcement – to the extent that the behavior that emerges involves a preference for punishing the guilty and protecting the innocent, this will emerge as a feature of the equilibrium rather than by assumption.
Intervention by a citizen on behalf of the authority or the target costs the untargeted citizen \( v \). If enforcement is “successful” (a concept developed more fully below), the target is sanctioned \( s \), whether or not she was a contributor. Let \( S \) be an indicator equal to one if enforcement is successful and zero otherwise. A fraction \( \lambda \in [0,1] \) of the sanction goes directly to the enforcer.

Finally, if citizen \( i \) has not been targeted, but some other citizen has, let \( A_i \) be an indicator equal to one if \( i \) attempts to intervene on behalf of the authority, and zero otherwise. Further, let \( T_i \) be an indicator equal to one if \( i \) attempts to intervene on behalf of the target, and zero otherwise. Note that such a citizen \( i \) may choose either to intervene on behalf of the authority or to intervene on behalf of the target, or to do neither. If the authority has chosen not to target any citizen, or if \( i \) himself has been targeted by the authority, both \( A_i \) and \( T_i \) equal zero.

The payoff to a citizen is derived from the endowment if kept, the public good, the sanction if imposed, and the cost of intervention if any (whether on behalf of the target or the authority).

\[
 u_i(y_c, r, s, v) = (1 - (1 - r)C_i)y_c + ry_c \sum_{j \neq i} C_j - E_iSs - (T_i + A_i)v \tag{1}
\]

The stage game payoff to the authority is derived from her endowment, the public good, the cost of enforcement, and the sanction:

\[
 u_a(y_c, r, s, e) = yg + ry_c \sum_{i=1}^{n} C_i + E(\lambda Ss - e). \tag{2}
\]

Next, let \( A \) represent the number of citizens who intervene on behalf of the authority conditional on an attempted enforcement, and \( T \) the number of citizens who intervene on behalf of the target, with \( A + T \in \{0,1,\ldots, n-1\} \). We assume that the probability of enforcement success is a linear function of the number of untargeted citizens who intervene on behalf of the authority and the target, and that in the absence of any intervention, the probability of success is \( \frac{1}{2} \). In particular,

\[
 \Pr(S = 1|A, T; n) = \frac{1}{2} + \frac{A - T}{2(n - 1)}, \tag{3}
\]

where the \( \frac{1}{2(n-1)} \) ensures that this probability is always bounded between zero and one. In this specification, if all untargeted players assist the authority, enforcement is successful with certainty, whereas if all untargeted players intervene on behalf of the target, enforcement is never successful.
2 Analysis of the One-Shot Game

We look for “non-perverse” symmetric weak perfect Bayesian equilibria. Weak perfect Bayesian equilibrium requires that (a) each player’s choices be sequentially rational given her beliefs at the time of choice and the other players’ strategies; and (b) beliefs about the other players’ types be consistent with prior beliefs, equilibrium strategies, and Bayes’ Rule on the path of play. Symmetry implies (a) all citizens play the same contribution and intervention strategies; and (b) given more than one similarly-situated enforcement target (e.g., two non-contributors), the authority targets each with equal probability. Finally, “non-perverse” entails focusing only on equilibria in which, when indifferent but given the choice, the authority will target a non-contributor over a contributor. Because, on the margin, targeting non-contributors deters non-contribution and targeting contributors encourages it, this can only improve the (ex ante) welfare of the authority, who, recall, benefits from public good provision. Practically speaking, this means that the authority will only ever target a contributor if all citizens contributed.

Citizen intervention. Irrespective of whether the citizens learn one another’s contribution choices, equilibrium behavior in the intervention phase of the one-shot game is trivial: because intervention either on behalf of an enforcement target or authority is costly and offers no prospective benefit, citizens should never intervene, either on behalf of the authority or their fellow citizen.

Enforcement. The prospective expected utility to the authority who targets no one is zero. Given the authority’s expectation of no citizen intervention post-targeting, the prospective expected utility to the authority who targets a non-contributor or contributor is identical and is given by

$$\mathbb{E}[u_a(E = 1)] = \frac{\lambda s}{2} - e.$$  

(4)

Enforcement will occur if this quantity exceeds zero. As noted above, we restrict attention to equilibrium strategy profiles in which the authority, conditional on preferring to target at all, randomizes over non-contributors, and targets a contributor only in the event that all citizens contribute.

Initial contributions. From the above, if a citizen contributes, he is only at risk of being targeted if everyone else contributes. Let $\theta^1$ represent the probability the authority targets a contributor conditional on all citizens contributing, and $\theta^0$ the probability she targets a non-contributor conditional on at least one citizen not contributing. Let $\kappa$ be the probability each citizen contributes. Recalling that similarly situated citizens are targeted with equal probability, the probability contributing citizen $i$ is successfully
targeted is then given by $\theta^1\kappa^{n-1}/2n$. Then the expected utility to the citizen of contributing is given by

$$E[u_i(C_i = 1)] = (\kappa + 1)ry_c - \frac{\theta^1\kappa^{n-1}s}{2n}. \quad (5)$$

The expected utility to a citizen of not contributing must take into account all of the states that might arise given different configurations of contribution choices by fellow citizens. This quantity is given by

$$E[u_i(C_i = 0)] = yc + \kappa ry_c - \frac{\theta^0s}{2} \sum_{j=0}^{n-1} \frac{1}{j+1} \binom{n-1}{j} \kappa^{n-1-j}(1 - \kappa)^j. \quad (6)$$

Note that if the authority’s best response is not to target ($\theta^0 = \theta^1 = 0$), then the expected utility to the citizen of contributing reduces to $(\kappa + 1)ry_c$, and the expected utility of not contributing to $yc + \kappa ry_c$. Comparing these quantities, non-contribution is a best response for all citizens.

By contrast, suppose the authority’s best response is to always target: ($\theta^0 = \theta^1 = 1$). If the sanction is sufficiently small, this will not induce compliance; likewise, if the sanction is sufficiently large, full compliance will be an equilibrium. More interesting cases emerge when sanctions are at an intermediate level: In such cases, multiple symmetric equilibria may exist: a no-compliance equilibrium, a full-compliance equilibrium, and a mixed strategy equilibrium in which citizens contribute probabilistically. The multiplicity of equilibria emerges due to strategic complementarities. The nature of these complementarities is intuitive: an authority can target one citizen at most, and weakly prefers targeting non-contributors to contributors. An no-contribution equilibrium may emerge because the risk of punishment is spread across the citizens: given that an individual’s fellow citizens are not contributing, a best response is also not to contribute. In the same setting, however, a full contribution equilibrium can also emerge: if all citizens are contributing, then a unilateral deviation of an individual to no contribution would bring about certain targeting. Thus the best response to full contribution by one’s fellow citizens is to contribute. In the mixed strategy equilibrium, citizens are indifferent between contributing and not contributing, and that indifference is sustained by the non-degenerate probability that each contributes.
Supplemental Appendix C.

Additional Information about Empirical Analysis

1 Means and Balance Statistics for Pre-Treatment Subject-Level Covariates

Table 1: Means and Balance Statistics for Pre-Treatment Subject-Level Covariates

<table>
<thead>
<tr>
<th></th>
<th>Appropriation</th>
<th>Salary</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.40</td>
<td>0.42</td>
<td>0.85</td>
</tr>
<tr>
<td>Age</td>
<td>20.0</td>
<td>21.0</td>
<td>0.34</td>
</tr>
<tr>
<td>&gt; 5 experiments</td>
<td>0.55</td>
<td>0.47</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Low Information</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.32</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>Age</td>
<td>20.0</td>
<td>21.0</td>
<td>0.26</td>
</tr>
<tr>
<td>&gt; 5 experiments</td>
<td>0.50</td>
<td>0.47</td>
<td>0.78</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Full Info</th>
<th>Low Info</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.42</td>
<td>0.40</td>
<td>0.86</td>
</tr>
<tr>
<td>Age</td>
<td>21.0</td>
<td>21.0</td>
<td>0.94</td>
</tr>
<tr>
<td>&gt; 5 experiments</td>
<td>0.47</td>
<td>0.47</td>
<td>0.95</td>
</tr>
</tbody>
</table>

|                      |               |         |         |
| Appropriation        |               |         |         |
| Male                 | 0.4           | 0.32    | 0.39    |
| Age                  | 20.0          | 20.0    | 0.52    |
| > 5 experiments       | 0.55          | 0.50    | 0.61    |

2 Description of Regression Specifications for Tables 2 and 4

In Table 2, we predict *net helping and hindering*, which is the number of group members who assist the authority minus the number who hinder the authority (-3 to +3) when the authority attempts enforcement. The model is estimated via OLS with standard errors clustered at the group level.
The fully specified regression model (column 2) takes this form:

\[ \text{Net helping and hindering} = \beta_0 + \beta_1 \text{Appropriation Treatment} + \beta_2 \text{Lagged Avg. Group Contributions} + \beta_3 \text{Lagged Avg. Resoluteness} + \beta_4 \text{Lagged Predatory or Perverse Targeting} + \gamma \times \text{Period} + e, \]

where the variables are defined as described in the main text and \( \text{Period} \) is a vector of period-specific indicators.

In Table 4, we simply add to this specification an indicator for the \textit{Limited Information} treatment in some specifications.

3 Discussion of Model Specification and Potential Threats to Inference

Sul (2013) presents a series of concerns about efforts to estimate treatment effects in repeated public goods experiments. In light of those concerns, we note that our statistical specification does not include a lagged dependent variable, which Sul shows may lead to bias if the lagged outcome is included, along with a treatment indicator, to predict contemporaneous outcomes. Instead, we predict \textit{Net helping and hindering} as a function of a treatment indicator, time period indicators, and a variety of (lagged) different behavioral measures. We present results with and without those covariates to highlight the potential role of those covariates in driving treatment estimates. Additionally, we supplement these statistical analyses both with graphical presentation of our data and simple summary measures of average behavior across treatment conditions.

Patterns of contributions in our experiments do not display strong time trends, in contrast to general trends in other public goods games that may lead to concerns about non-stationary time series estimates. These concerns are larger in models that include lagged contribution behavior, which we do not include in our specifications. (Additionally, we instead include period indicators to account for any pattern of period-specific contribution behavior.)

References