

Identifying Legitimacy: Experimental Evidence on Compliance with Authority*

Eric S. Dickson Sanford C. Gordon Gregory A. Huber

Abstract

We consider the extent to which individuals' perceptions of an authority's legitimacy affect their intrinsic motivation to comply with that authority. Studying legitimacy empirically is challenging because actions or institutions that might affect intrinsic compliance motivations typically affect extrinsic material motivations as well. We propose a public goods experiment that disentangles these motivations by manipulating the extrinsic channel probabilistically. Costly action by an authority to improve the fairness of her enforcement efforts significantly improves pro-social behavior even when the extrinsic channel is held fixed. A follow-up experiment demonstrates that this improvement is not motivated by a desire among citizens to "pay back" authorities who took pains to improve fairness, but is consistent with a mechanism of normative legitimation.

*Word count: 9,799

1 Introduction

This paper considers how an individual’s decision to comply with rules and behavioral norms is shaped by perceptions of an authority’s legitimacy. The normative question of what constitutes legitimate authority has preoccupied political and legal philosophers at least since Plato, but concern regarding its empirical antecedents spans the social sciences. Because unlimited coercive capacity is prohibitively costly or otherwise infeasible, governing authorities who wish to facilitate social order must rely at least in part on a widespread consensus that complying with their edicts is “the right thing to do” (North, 1981; Levi, 1988).

Beyond that, however, social scientists, philosophers, and policymakers care about legitimate authority because they care about fairness, with the understanding that perceptions of unfairness can undermine the social compact. A series of recent incidents around the United States involving aggressive police tactics, particularly in communities of color, have undermined trust in the police (Pew Research Center, 2016) and brought these issues to the forefront of national political consciousness. An improved understanding of how authorities can effectively enforce the law on the one hand, while simultaneously treating citizens fairly and maintaining legitimacy on the other hand, could yield substantial benefits for citizen well-being.

To construct an analytically-grounded account of the antecedents of legitimate authority, we begin with a definition from social psychology: Tyler (1997) defines legitimacy as a “judgment by group members that they ought to voluntarily obey social rules and authorities *irrespective of the likelihood of reward or punishment*” (emphasis ours). In other words, legitimacy obtains when the features of authoritative institutions, or the choices of individuals in positions of authority, enhance the *intrinsic* motivations of citizens to carry out certain social duties.

Empirically identifying how an authority’s actions can affect these intrinsic motivations is challenging because of a fundamental problem of causal attribution. The problem arises because procedures or choices that enhance these motivations may simultaneously alter cit-

izens’ beliefs about the authority’s capacity to bestow rewards and punishments. These extrinsic factors may, in turn, affect citizen behavior, quite apart from considerations of the authority’s legitimacy.¹ This attribution problem is particularly pronounced in the realm of policing, where, for example, allegations of abuse invariably call into question not only the efficacy, but also the fairness – and thus legitimacy – of law enforcement.

To sharpen the point somewhat, suppose a police officer behaves in a visibly “fair” manner, being scrupulous in her attention to detail and courteous to the citizens with whom she interacts. If citizens on her beat obey the law more than they do a few blocks over, is it because they perceive her as legitimate, or because they perceive her as competent and thus less error-prone? If her actions reduce the likelihood that she makes mistakes (arresting the innocent or failing to arrest the guilty), that will improve compliance through citizens’ extrinsic motivations. But the relationship between her fair actions and citizens’ compliance may also be driven by a positive affect toward the officer herself and the institutions she represents, which enhances citizens’ intrinsic motivations to obey the law.

Or, consider the following example from the Department of Justice report on the Ferguson, Missouri police force (Department of Justice, 2015). According to the report, “City, police, and court officials for years have worked in concert to maximize revenue at every stage of the enforcement process...” (p. 10). This, when combined with extant racial disparities in policing, severely degraded the citizenry’s trust in the police. Now suppose that citizens in Ferguson, as a consequence of mistreatment, are less likely to comply with the law’s edicts than those in a similar town with a different law enforcement culture. A legitimacy-based account attributes this to the fact that citizens in Ferguson feel no psychological predisposition toward, or attachment to, the authority. But a deterrence-based account might attribute relative noncompliance to the fact that the police in Ferguson were less interested in appropriately punishing the guilty than they were in punishing as many people as possible,

¹For a related critique of the analytical utility of the concept of legitimacy in the sociology of law, see Hyde (1983).

thereby undermining material motives to comply.

To circumvent the attribution problem, we present an experimental paradigm in which the effects of an authority’s behavior on the material incentives of citizens are carefully delineated and decoupled from its effect on intrinsic ones through a novel randomization. In our framework, an authority can take a costly action meant to improve the fairness with which failures to contribute to a public good are punished. However, the action is only effective with a known and predetermined probability. Citizens are made aware both of the authority’s decision to engage in the action as well as the actual material incentives they face, based on the randomly realized degree of fairness of the institution. We are thus able to measure directly the effects of the authority’s action per se on citizen behavior, holding constant the citizens’ material incentives.

Our paradigm, which employs a public goods contributions framework, relates to experimental research on cooperation with third party punishment in different institutional contexts. Closest to our approach are papers by Dal Bó, Foster, and Putterman (2010), Baldassarri and Grossman (2011), and Grossman and Baldassarri (2012), who adopt an experimental approach to consider whether electing a centralized sanctioning authority can improve cooperation. Our experiments permit us to isolate a legitimating effect induced by a central authority’s costly attempt to improve an enforcement mechanism’s accuracy in assigning punishments to non-contributors (i.e., its procedural fairness).

In our main experiment, we find strong evidence that the authority’s fairness-seeking action markedly increases citizens’ intrinsic motivations to engage in pro-social behavior, independent of the material consequences of that action. Specifically, the authority’s mere attempt to implement a fairer procedure increases the probability that a citizen contributes to the public good by 10 to 12 percentage points.

The effect we observe in the main experiment is consistent with several potential mechanisms: (1) a “normative legitimation” account in which the authority’s action conveys information about her own quality that independently primes a citizens’ sense of obligation

to behave in a pro-social manner; (2) a “conditional cooperation” account in which the authority’s action conveys information about *other citizens*’ willingness to contribute; and (3) a “direct reciprocity” account in which the citizen seeks to “pay back” the authority for the costly investment.

Analysis of ancillary results from our main experiment cast doubt on the second mechanism. In a follow-up experiment, we consider whether the results from the main experiment are driven by reciprocity. The experiment accomplishes this by randomizing whether or not the authority materially benefits from the public good. We find that the nature of the authority’s compensation does not moderate the effect of the authority’s action. While we cannot eliminate all conceivable alternative mechanisms, this conjunction of findings establishes the preeminence of the normative legitimation account.

The question of what makes authority legitimate has occupied empirical social scientists since Weber’s (1978) seminal writings on the subject in the early 20th century. Recently, the prominent behavioral research on the subject has emerged at the intersection of social psychology (e.g., Tyler, 1990; Tyler and Ho, 2002) and criminal justice (e.g., Bottoms and Tankebe, 2012), with particular attention paid to the relationship between perceptions of procedural justice (Lind and Tyler, 1988) and legitimacy. Political science research on legitimate authority has paid closest attention to attitudes on legitimacy and the closely related notion of “diffuse support” (Easton, 1965), with a particular focus on support for courts (e.g., Caldeira and Gibson, 1992; Gibson, Caldeira, and Baird, 1998) and international organizations (e.g., Hurd, 1999). A common thread in these two political science research traditions is the question of how institutions with limited enforcement power achieve deference in spite of their politically tenuous position.

One challenge each of these literatures faces is the difficulty of isolating legitimacy-based from other, possibly extrinsic, motivations to support institutions. Thus, one contribution of our research is to provide an analytical framework and empirical approach for disentangling extrinsic and intrinsic motivations to comply with authoritative institutions, and also to

decompose the sources of those intrinsic motivations. In the conclusion, we discuss the broader implications of isolating these effects for institutional and policy design.

2 Conceptual Issues

2.1 A Fundamental Problem of Identification

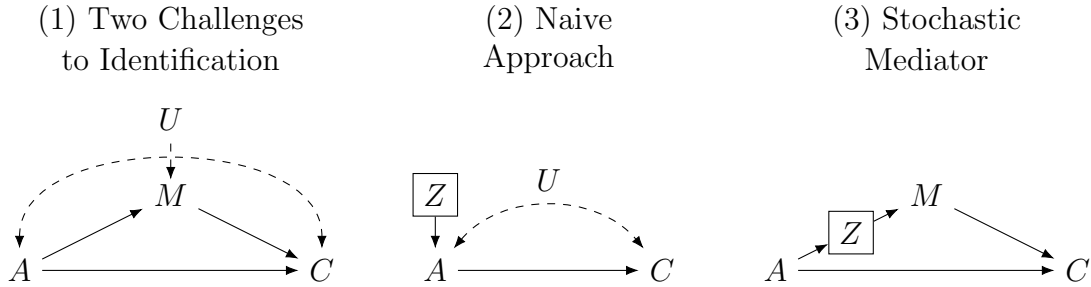
In the social science literature, a standard approach to estimating the legitimating effect of procedurally fair institution or authority action is survey-based (e.g., Tyler, 1990; Paternoster et al., 1997; Murphy, 2005): query citizen i who has had an encounter with the law or another authority about whether or not they were treated fairly ($A_i \in \{unfair, fair\}$) and whether or not they intend to comply with the law in the future ($C_i \in \{don't\ comply, comply\}$). Such studies interpret an association between variables A and C as evidence that it is subjects' perceptions of the legitimacy of the legal process that affect their compliance intentions.

Panel (1) of Figure 1 depicts a directed acyclic graph (DAG) intended to illustrate two inferential issues with this approach. The first is omitted variables bias: as noted by van den Bos (2001), perceptions of fairness may be correlated with numerous unobserved individual- and institutional-level factors that also explain compliance (denoted U in the figure).² The other inferential challenge is that perceptions of fairness may affect compliance via the two channels described in the introduction: an extrinsic channel in which A affects material motivations M , which in turn affect C ; and an intrinsic channel in which A 's actions directly affect C . Even absent unobservable confounds, an observed association between A and C cannot distinguish between these channels. Studies that randomize the fairness of an experimental treatment (e.g., Vermunt et al., 1996; van den Bos, 2001) may solve the issue of unobservable confounds, but remain susceptible to this attribution problem.³

²Measurement error also poses a threat to inference: citizens may misreport their perceptions or behavior, and even random measurement error of a control may artificially inflate the coefficient on the treatment of interest.

³Note that in this research the measured outcome is affect toward an institution, not

Figure 1: Directed Acyclic Graphs Demonstrating Challenges to Identifying Legitimizing Effects of Authority Choices, and a Potential Solution



Panel (1) depicts two inferential challenges with uncovering the legitimating effects of an authority’s actions. Panel (2) depicts an experimental approach that assumes away the attribution problem. Panel (3) depicts our stochastic mediation approach for circumventing the attribution problem. A denotes the authority’s action; M material incentives; C the citizen’s compliance choice; U unobserved confounders; and Z a random assignment variable.

Given these challenges, an alternative design would randomize some treatment thought to proxy a legitimating action by an authority, and restrict attention to treatments that have no effect on the citizens’ material incentives. This “naive” approach, depicted in panel (2) of the figure (with the random variable Z determining treatment assignment), can recover valid estimates of the average treatment effect of A on C (either directly, if Z perfectly determines A , or via instrumental variables estimation otherwise), but introduces two problems of ecological validity. First, it precludes studying any legitimating action related to the fairness of payoff-relevant factors in a citizens’ compliance choice. Insofar as fairness is nearly always defined with respect to materially relevant outcomes such as accuracy or bias in the application of punishment, this domain restriction is quite severe. Second, randomization undermines the ability of citizens to form beliefs about the authority herself – beliefs that surely form the foundation of any account of her legitimacy.⁴

Our experimental approach to circumventing these challenges draws on recent advances in compliance.

⁴In the extreme, a person compelled to take an action does not signal anything about herself by that choice.

in design-based approaches to identifying causal mediation effects (Imai, Tingley, and Yamamoto, 2013; Pirlott and MacKinnon, 2016) and is depicted in panel (3). First, note the absence of the bidirected arc corresponding to unobservable confounds U between A and C . We shut down this path via randomly matching subjects assigned the citizen and authority roles. Second, we sever the deterministic link between the authority’s legitimating action and the citizen’s material incentives by introducing a stochastic mediating variable (Z) triggered by the authority’s action, which in turn determines the level of the material incentives M .

Suppose that Z is more likely to take on values that lead to high levels of M when $A = 1$ than when $A = 0$. (Our main experiment is slightly more involved, but the intuition carries over.) If we condition on a value of $Z = z$ that is feasible both when $A = 1$ and $A = 0$, comparing values of C under those different values of A identifies the direct effect of A on C given $Z = z$. Conditioning on A and comparing values of C under different levels of M , by contrast, identifies the conditional causal mediation effect of material incentives on citizen behavior in either the presence ($A = 1$) or absence ($A = 0$) of the authority’s action.

Informally, then, this design permits us to identify the effect of an authority’s “good faith effort” to improve the fairness of her enforcement work on citizens’ subsequent compliance decisions. Further, the design permits a comparison of the magnitude of this “legitimizing” effect to the effect of changes in material incentives on compliance decisions. That being said, the design as stated does not permit us to decompose the various psychological mechanisms that may contribute to the legitimating effect – an issue that animates our second experiment, described in greater detail below.

2.2 Possible Mechanisms Underlying the Intrinsic Channel

While the identification strategy described in the preceding section permits us to distinguish between the extrinsic and intrinsic channels, it does not, by itself, distinguish among the various psychological mechanisms that might animate the latter. Here, we consider three.⁵

⁵A fourth, which involves the authority’s action as a coordination device (cf., Myerson, 2004), is ruled out in our experiments by the absence of strategic complementarities among

Normative Legitimation. The first mechanism we consider comes closest to the idea of legitimate authority described in the social scientific literature cited above as well as by normative political theorists (e.g., Anscombe, 1981; Habermas, 1979; Raz, 2009). Underpinning the normative account is the notion that an authority’s legitimacy depends on the view of its subjects that it is worthy of recognition as such, and that this recognition entails some (finite) sense of obligation to comply with the authority’s edicts. Behaviorally, the normative legitimation mechanism would involve citizens updating their beliefs about the quality of the authority herself from the costly action, with the resultant sense of esteem fostering an intrinsic motivation to comply.

Conditional Cooperation. According to the second mechanism, the authority’s behavior conveys information to citizens about the prevalence of pro-social types among *other citizens*, with greater regard for those citizens inducing stronger motivation to contribute out of an anticipation that those citizens will, too, contribute (e.g., Fischbacher, Gächter, and Fehr, 2001). This mechanism is plausible in our experimental setting because authorities and citizens are drawn from the same population of subjects. It also implies several ancillary empirical implications, which we test below.

Direct Reciprocity to the Authority. A third mechanism involves a direct form of reciprocity between citizens and authorities. By choosing a costly action that improves the fairness of enforcement, the authority might be seen as having done right by citizens. Citizens, may in turn feel intrinsically motivated to “pay back” the authority for having undertaken this investment (cf., Akerlof, 1982). In our first experiment, citizens could achieve this goal by contributing to the public good, because the authority herself benefits from these contributions. Our second experiment is explicitly designed to distinguish the direct reciprocity mechanism.

We do not mean to suggest that this constitutes an exhaustive accounting of all conceivable citizens: the payoff relevant consequences of a citizen’s choice is unaffected by the choices of other citizens.

able mechanisms; and there is also little reason to believe *ex ante* that only one mechanism is relevant on the broad topic of the relationship between authorities and citizens. That said, our analysis below provides an approach to adjudicating among these mechanisms, while also suggesting how one might go about testing, in future work, additional mechanisms not described here.

3 Study 1: Authority Investment in Fair Enforcement

3.1 Design

We now elaborate on a specific instantiation of the design described above. We employ an experimental public goods setting, in keeping with the longstanding practice in experimental social science of using contributions in a public goods game as a metaphor for social norm compliance (Ledyard, 1995).

Specifically, the setting of the experiment is a linear public goods game with a centralized authority: subjects choose whether or not to contribute to a public good and are subject to enforcement actions by another subject designated as an authority. The authority and her “citizens” share a common interest in maximizing contributions to the public good. (This is analogous to a situation in which a police officer lives in the community he or she polices.) Although individual citizens would prefer to withhold their contributions irrespective of the behavior of other citizens, citizens would nonetheless be better off if all players contributed than if none did so.

Subjects interacted anonymously via networked computers. The experiments were programmed and conducted using the software *z-Tree* (Fischbacher, 2007). After giving informed consent according to human subjects protocols, subjects received written instructions that were subsequently read aloud to promote understanding and induce common knowledge of the experimental scenario. No deception was employed. Before beginning the experiment, subjects took an on-screen quiz that both measured and promoted understanding of the instructions.

Subjects earned tokens, convertible into dollars at the end of the experiment (30 tokens = US\$1) in amounts determined by the outcomes of play. Subjects' overall payoffs in a given session were equal to the sum of payoffs from each of the 20 periods (converted into dollars), plus a US\$7 show-up fee.

At the beginning of each period, subjects were each given an endowment of 20 tokens and randomly assigned to a group of five people, of which four were randomly assigned as citizens (Role A, in the neutral parlance of the experiment), and one as an authority (Role B). Group and role assignments were randomly reassigned after each period. In each period, individual group members in Role A were labeled with an ID number between 1 and 4, commonly known to be randomly assigned in each period. Each period consisted of one play of the following extensive form game:

1. Authority chooses to make a “big” (4 token) or “small” (0 token) investment in accuracy.
2. Accuracy level given authority's investment determined.
3. Authority learns realized accuracy level and chooses enforcement rule.
4. Each citizen learns authority's investment, realized accuracy level, and enforcement rule, and chooses whether or not to contribute endowment to common pot.
5. Signals generated and enforcement rule implemented; payoffs realized.

We now turn to a fuller description of key features of the design.

Accuracy. Depending on its authority's choice, each group would be assigned to one of three different accuracy levels. (1) Under “Low Accuracy Information,” the signals generated about each individual citizen's contribution decision has a 40% error rate. This means that if a specific citizen in fact kept (allocated) his tokens, the computer would generate a signal that the citizen kept (allocated) his tokens with 60% probability, but would generate an incorrect signal that the citizen allocated (kept) his tokens with 40% probability. (2) Under “Medium

Accuracy Information,” the error rate is 25%. (3) Under “High Accuracy Information,” the error rate is 10%.

If the authority chose the small investment, the group’s realized level of accuracy would be low with 50% probability and medium with 50% probability. If the authority chose the big investment, the realized level of accuracy in the group would be medium with 50% probability and high with 50% probability. The realized level of accuracy is revealed to the authority before her choice of enforcement rule, and to the citizens (along with the authority’s investment choice and the enforcement rule) prior to their contribution choice.

The decision to undertake the big investment is our instantiation of a legitimating action, because it leads to an enforcement institution that is procedurally more fair.

Enforcement Rule. The enforcement rule chosen by the authority in stage 3 is carried out automatically in stage 5, and thus represents a binding commitment revealed to the citizens prior to their contribution choice. Authorities could select one of four enforcement rules:

- Deduct 24 tokens from each citizen for whom a signal of “kept” was generated
- Deduct 24 tokens from each citizen for whom a signal of “allocated” was generated
- Never deduct tokens from any citizen, irrespective of signals generated
- Deduct 24 tokens from all citizens, irrespective of signals generated

As a shorthand, we will refer throughout to the first of these rules as “PATS” (Punish According To Signal); the second as “anti-PATS;” the third as “never punish;” and the fourth as “always punish.”

We will focus much of our attention on the large majority of cases in which the authority chose the PATS enforcement rule, a decision we explain in greater detail below. The function of the enforcement rule choice is to create conditions under which the subjects assigned to the role of citizen mentally associate the authority with the dispensing of punishment.

Precommitment greatly simplifies the strategic problem for the citizens, by creating common knowledge about the administration of penalties. In the absence of precommitment, citizens might condition their contribution choices on posterior beliefs about the resoluteness of authorities based on the accuracy investment, in addition to higher order beliefs about the beliefs of other citizens.

Extrinsic Incentives. The marginal per capita rate of return (MPCR) for citizens and authorities alike is 0.4, meaning that for every 20-token contribution to the common pot, the authority and each citizen receive 8 tokens. The 24-token deduction described above is calibrated to make a citizen motivated purely by the material payoffs of the experiment indifferent between contributing and not contributing under medium accuracy and PATS. To see this, note that a citizen who does not contribute under PATS/Medium Accuracy keeps his or her 20 token endowment but has 24 tokens deducted as a punishment with .75 probability, yielding an expected payoff of $20 - 0.75 \times 24 + \hat{C}_{-i} = 2 + \hat{C}_{-i}$ tokens, where \hat{C}_{-i} represents citizen i 's beliefs about others' contributions. At the same time, a citizen who does contribute receives a return of $0.4 \times 20 = 8$ tokens from his own contribution, but has 24 tokens deducted as a punishment with only 0.25 probability, yielding an identical expected payoff of $8 - 0.25 \times 24 + \hat{C}_{-i} = 2 + \hat{C}_{-i}$ tokens. As described below, our identification strategy focuses on comparisons within the Medium Accuracy institution given the PATS enforcement rule. The material indifference these parameter values induce is therefore useful because it should, a priori, maximize variation in citizen contributions in that setting, which will be critical for our identification goals. Very strong or very weak material incentives to contribute would, by contrast, obscure the possibility that legitimating actions per se might have a causal effect on citizen behavior.

Citizens' incentives to contribute to the public good are higher under PATS than under alternative enforcement strategies, even in the presence of imperfect signals about the contribution decision. Accordingly, an authority motivated by marginal deterrence alone is

always weakly better off selecting this enforcement rule.⁶ Additionally, PATS is the only enforcement rule for which more accurate information improves the material incentives to comply. For these reasons, we focus the bulk of our analysis below on cases in which the authority chooses PATS (and thus holding constant the enforcement rule).

Finally, because players are randomly assigned to new groups and roles at the end of each period and interact anonymously, they have no reason to condition their choices on behavior in past rounds, or in expectation of future actions or repeated interactions (e.g., cultivating reciprocity norms).

3.2 Identification

At the moment citizens are choosing whether to contribute to the public good, they are fully informed about all factors affecting their extrinsic motivations to comply: all parameter values, the realized level of accuracy, and the enforcement rule. (Recall that choices of other citizens drop out of the material payoffs calculation.) Additionally, they have also observed whether the authority made the big investment in accuracy. Conditional on the realized accuracy level, this choice by the authority is materially irrelevant. However, it may not be irrelevant to citizens' intrinsic incentives to comply.

We are now in a position to make the experiment's instantiation of the identification strategy described in Section 2 explicit. Assume the authority has chosen the PATS enforcement rule. Citizens will then make their decisions in one of four circumstances, summarized in Table 1. Rows denote the authority's investment (Small or Big), while columns denote the realized accuracy level. Conditional on the authority's choice of Small Investment, the citizen is randomly assigned to Low or Medium accuracy. Conditional on the authority's choice of big investment, the random assignment is to Medium or High investment. The notation $\bar{C}_{r,k}$ denotes sample average group-level contribution levels in row r and column k .

⁶In the presence of idiosyncratic shocks to citizens' motivations to contribute, PATS is the only enforcement rule consistent with equilibrium play.

Table 1: Identification of Legitimacy Effect, Holding Punishment Strategy Constant

		Realized Accuracy Level		
		Low	Medium	High
Authority Investment	Small	$\bar{C}_{S,L}$	$\bar{C}_{S,M}$	—
	Big	—	$\bar{C}_{B,M}$	$\bar{C}_{B,H}$

Cell entries denote sample average group contribution rates

An approach analogous to one conducted in much of the prior literature would be to pool rows and compare average contribution rates given the big and small investments. It is immediate, however, that this confounds extrinsic and intrinsic motivations to contribute. The more appropriate comparison of interest is

$$\bar{C}_{B,M} - \bar{C}_{S,M},$$

the differences in contributions given Medium Accuracy between an authority who makes the big and small investment. Given the PATS enforcement rule, this holds all material motivations for citizens to contribute constant. The only difference between the two circumstances is that in one, the authority undertook a costly action to improve the institutional environment, while in the other she did not. Thus, this comparison yields a valid causal estimate of the effects of an authority's legitimating action per se.

We will also consider two institutional comparisons. The first is $\bar{C}_{S,M} - \bar{C}_{S,L}$. This identifies the effect of a change from Low to Medium accuracy, holding the authority's investment choice constant at Small. The second is $\bar{C}_{B,H} - \bar{C}_{B,M}$, which identifies the effect of a change from Medium to High accuracy, holding the authority's choice constant at Big. Based on the logic of the previous section, each of these differences identifies the effect of the realized accuracy level (for a given action by the authority).

Of course, there are other factors in the experiment that may also affect citizen behavior. Individuals' experiences during earlier rounds may affect their later play, and their behavior may generally evolve over the course of the game. While our design in which individuals are randomized into different groups and roles over time and then interact anonymously seeks to prevent most sources of repeat-play dynamics, we nonetheless undertake a variety of approaches to account for such dynamics in the analysis that follows.⁷

3.3 Results

We conducted four experimental sessions at [Redacted] and two sessions at [Redacted]. Each of the 90 subjects who participated took part in one session only. At both institutions, participants signed up via a web-based recruitment system that draws on a large, pre-existing pool of potential student subjects. (Subjects were not recruited from the authors' courses, and did not receive course credit for participating.) 46% of the subjects were female, and the median age was 20. 8% of the subjects were Economics majors, though 27% majored in a social science department. Subjects earned an average of \$22.91 (s.d. of \$2.5), with a maximum of \$30.10 and a minimum of \$17.70. The average score on the quiz administered between the reading of the instructions and play of the experiment was 6.4 out of 8 (s.d. 1.6) with 60% of subjects receiving a score of 7 or higher, and 24% receiving a perfect score.

Authority Behavior. The data consist of 360 group-period interactions. Table 2 summarizes the authorities' investment and enforcement rule choices in those interactions. Authorities selected the "big" investment 214 out of 360 times, or 60%. There was a modest increase in investment over time: in the first five periods the average rate was 50%, and in the last five it was 64%. Among players who were in the authority role more than once, 41% choose each investment level at least once. As a consequence of these choices, in 21% of all group-periods the authority received low accuracy information, in 53% medium accuracy

⁷Even if players condition their behavior on their prior group-level experiences, groups are reshuffled across rounds so those expectations should be identical across groups in subsequent rounds.

Table 2: Authority Choices in the Baseline Experiment: Group-Level Data

	Enforcement Rule				Total
	PATS	Anti-PATS	Never Punish	Always Punish	
Small Investment	81	7	40	18	146
Big Investment	185	8	15	6	214
Total	266	15	55	24	360

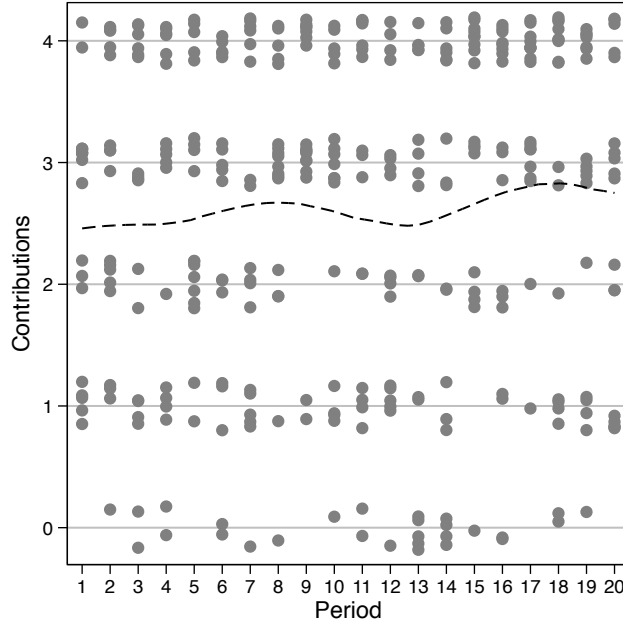
information, and in 27% high accuracy information. 97% of players experienced all three accuracy levels while in the citizen role and the remaining 3% experienced two.

Those acting in the authority role overwhelmingly chose the PATS (punish according to signal) enforcement rule (74% of the time), although a substantial minority chose to never punish (15%), and smaller proportions chose either to always punish (7%) or to punish according to the anti-PATS rule (4%).⁸ The PATS rule is slightly less common in the first 5 periods of play than afterwards: PATS is chosen by 67% of players in the first five rounds compared to 76% of the time in the remaining periods. Those who made the big investment in accuracy are more likely to choose PATS than those who made the small investment (87 versus 55% of the time), with those who made the small investment more likely to choose either to never (27 versus 7%) or always (12 versus 3%) punish. Because the investment decision affects the accuracy of the signals received by the authority, there is a similar relationship between realized accuracy levels and the enforcement rule.

Aggregate Citizen Contribution Behavior. Overall, citizens contributed their tokens to the public good 65% of the time, for an average group contribution rate of 2.6 out of 4 (median 3). Unlike in standard public goods games with no enforcement (e.g., Fehr and Gächter, 2000), contributions do not diminish over time. Figure 2 displays data on group-level contributions by period (data points are jittered for clarity), along with a local polynomial smoother. The average contribution rate rises slightly over time, from around

⁸Two players accounted for 47% of the cases of anti-PATS, while eight players chose the anti-PATS rule only once.

Figure 2: Group Contributions by Period

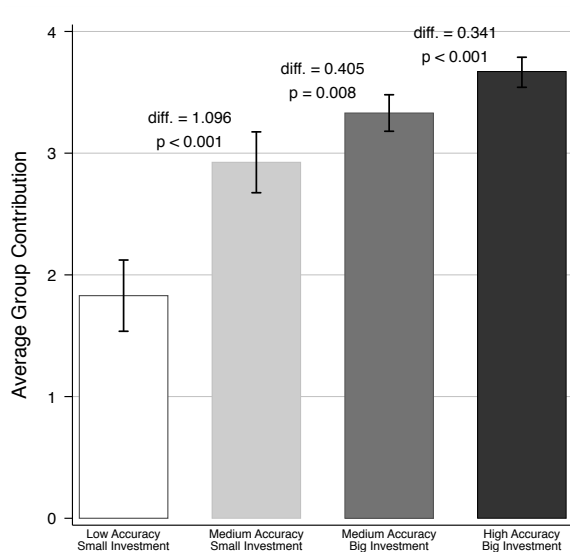


Note: Data jittered to enhance clarity of presentation. Dashed line is local polynomial smoother.

2.5 in the first five periods to 2.8 in the final five. 94% of players varied their contribution decisions, while the remaining players nearly evenly split between never and always contributing.

The figure masks considerable heterogeneity in the data, which we explore systematically below. Unsurprisingly, contribution levels are highest when the authority adopts the PATS enforcement rule. Under PATS, the average group contribution rate is 3.1 out of 4; under Never Punish, 1.0; under Always Punish, 1.4; and under anti-PATS, 0.8. (The median rate was 3 out of 4 under PATS and 1 out of 4 under the remaining enforcement rules.) The stark difference between citizen behavior under PATS and under the other enforcement rules is likely due in large part to the much greater likelihood of enforcement errors: aggregating over different accuracy levels, under PATS, non-contributors escaped punishment 32% of the time, and contributors were punished 22% of the time. By contrast, under all other enforcement rules and accuracy levels, non-contributors escaped punishment 72% of the time, while the

Figure 3: Group-level Contributions by Investment Decision and Accuracy Levels Given the Authority Choses Punish According to Signal



Note: Averages with bootstrapped 95% confidence intervals.

corresponding figure for punishment of contributors was 42%.

Institutional Effects. The experiment permits us to identify two institutional effects: the effect of a change from low to medium accuracy conditional on a small investment by the authority (and the PATS enforcement rule), and the effect of a change from medium to high accuracy conditional on a big investment (and PATS). The relevant data are displayed graphically in Figure 3, which plots group average contributions by the authority's investment and realized accuracy level, conditional on the PATS enforcement rule. The first institutional effect is obtained by comparing the pale gray column (bar #2, medium accuracy, small investment) with the white one (bar #1, low accuracy, small investment). The average group-level contribution rate in the former category was 2.93, as compared with 1.83 in the latter. The difference of 1.1 contributions is highly statistically significant and implies that the change in accuracy from low to medium induces a 60% increase in contributions.

The second institutional effect is retrieved by comparing the black column (bar #4, high accuracy, big investment) with the dark gray one (bar #3, medium accuracy, big investment).

Here, the effect is smaller, but still statistically significant: moving from medium to high accuracy conditional on the big investment is associated with an increase in group-level average contributions from 3.33 to 3.67, an increase of 10.2%.

Effect of the Legitimizing Action. We turn next to our analysis of the direct effect of the authority’s action, which is calculated as the different in contribution rates given the authority’s big and small investment, holding constant accuracy (at medium) and enforcement rule (at PATS). Turning first to group-level average contributions, this calculation can be made comparing the dark and light gray bars (bar #3 and bar #2, respectively) in Figure 3. Under medium accuracy and a small investment, the average group-level contribution is 2.93 out of four. Under medium accuracy and a big investment, the group-level contribution is 3.33 on average. The difference in means, about 0.41, is significant at $p < 0.01$ (two-tailed), and corresponds to a 14% increase in the contribution rate.

Using simple raw descriptions of the data and leveraging the design features of our experiment, we have provided direct evidence of a legitimacy effect: authorities who attempt to improve the quality of information on which they condition their punishments induce greater contribution levels by their group members than those who don’t, holding the realized quality of information constant. An authority’s legitimating action therefore appears to be associated with real changes in citizen behavior.

Robustness. For the next part of the analysis, we allow for individual-level covariates and effects, shifting from a focus on group-level outcomes to the decisions of individual players in the citizen role. Specifically, we model individual i ’s decision whether to contribute in period t , $C_i \in \{0, 1\}$, as a function of the treatment variable (the authority’s choice of the big investment in i ’s group g in time t , $a_{g,t} \in \{0, 1\}$), and additional covariates. Data are restricted to cases in which the authority chose the PATS enforcement rule and realized accuracy is medium. All specifications are OLS regressions with standard errors clustered at the group-period level (i.e., a group of four citizens in a single group in a single period),

the level at which treatment is applied. Our baseline specification, then, is simply

$$C_{i,t} = \beta_0 + \beta_1 a_{g,t} + \gamma X_{i,t} + \varepsilon_{g,t}.$$

This framework allows us to understand whether our results are affected by accounting for players' experiences earlier in the game, the period of play, or other factors.

Table 3 shows these results. Note that because there are four citizens in a group, we should, in the absence of serious confounding, expect treatment effects estimates for the individual-level contribution analysis to be about 1/4 the 0.41 group-level difference in means described above. Column (1) is a simple OLS specification that mechanically demonstrates this: the effect of the big investment in this specification is around 10 percentage points ($p < 0.01$, two-tailed).⁹ In column (2), we include period indicators, which does not materially affect the estimated result. Column (3) adds each player's average experienced group contribution rate prior to the current period. Players who have experienced groups with more contributions in the past are more likely to contribute, but accounting for this effect modestly increases our estimate of the legitimating effect of the authority's investment in the current period. In the column (4) specification, we also account for each player's experience of how frequently the authority in prior periods made the big investment. The estimated legitimacy effect remains positive and statistically significant.

In the column (5) specification, we restrict our analysis to players who had already served as the authority in at least one prior period (recall that subjects are randomly reassigned to new roles in each period), in case the absence of such prior experience meaningfully affects the way in which subjects understand the implications of the investment choice. In this specification, we continue to find that individuals contribute more to the public good when

⁹As noted in the text, this specification clusters at the period-group level. Both unclustered and robust standard errors are smaller. If we instead cluster at the session level, standard errors are slightly larger, with $p = .06$, two-tailed.

Table 3: Estimated Effect of Authority's Investment Choice, Given Medium Accuracy and PATS Enforcement Rule

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Authority Investment	0.101	0.109	0.115	0.125	0.118	0.111	0.104	0.124
(1=big, 0=small)	[0.038]	[0.037]	[0.037]	[0.038]	[0.048]	[0.052]	[0.039]	[0.046]
Average group contributions			0.072					0.078
prior to this period (0-4)			[0.030]					[0.064]
Average investment experience				-0.083				-0.047
prior to this period (0-1)				[0.081]				[0.144]
Constant	0.731	0.635	0.588	0.768	0.929	0.770	0.747	0.597
	[0.032]	[0.063]	[0.105]	[0.098]	[0.039]	[0.063]	[0.025]	[0.139]
Observations	560	560	540	540	404	338	447	428
R-squared	0.01	0.04	0.05	0.04	0.05	0.08	0.02	0.07
Number of unique subjects							65	65
Period fixed effects	N	Y	Y	Y	Y	Y	N	Y
Subject fixed effects	N	N	N	N	N	N	Y	Y

Dependent variable is player contribution decision (1=yes, 0=no).

OLS coefficients with group-period clustered standard errors in brackets.

Observations are individual contribution decisions given medium accuracy and PATS enforcement strategy.

the authority chooses a big investment. To reduce the possibility that our results are due to the behavior of subjects who did not understand the formal structure of the experiment, in column (6) we restrict our analysis to players who got at least 7 of the 8 quiz questions measuring subject comprehension correct prior to the beginning of the experiment.

Our most conservative analysis appears in columns (7) and (8). In these specifications, we include individual-level fixed effects. Estimating these models entails restricting the sample to players who experience medium accuracy and PATS in the citizen role following both a small and big investment by an authority. In the specification that does not include controls for period or past experience the estimate is 0.114 ($p < .01$). In the specification with these controls, it is 0.124 ($p < .01$). Thus, even after accounting for each player’s individual propensity to contribute, we continue to find evidence that the authority’s legitimating choice directly increased the likelihood a player contributed to the public good.

3.4 Interpreting Study 1: Potential Mechanisms

Study 1 shows that citizens’ intrinsic incentives to engage in prosocial behavior can be markedly increased when an authority takes a costly action to improve the fairness of enforcement. It does not, however, adjudicate among the the three mechanisms described above. Study 2, described below, is specifically intended to isolate the efficacy of the direct reciprocity mechanism. Here, we discuss two indirect tests of the conditional cooperation mechanism described above.

If this mechanism were the primary driver of our Study 1 results, we would anticipate, first, that conditioning on the citizen’s prior experience with group contributions and authority behavior would attenuate the estimate of the intrinsic effect. The data do not bear out this hypothesis, however. Specifically, columns (3) and (4) of Table 3 include direct measures of each citizen’s prior experience, a proxy for each player’s beliefs about the behavior of others. Doing so does not reduce the estimated effect of interest.

Additionally, although the authority’s choice may provide novel information about other citizen’s proclivities, we would expect it to be most influential in earlier rounds, when players

have less experience with other players' contribution behavior. Each period provides three pieces of evidence about those tendencies (the contribution choices of three fellow citizens in the group), compared to only a single authority investment choice. For this reason, it is notable that we find that our estimate for the intrinsic channel is robust and stable across subsets of the periods of the game. Figure 4 displays average contribution rates under medium accuracy and PATS, conditional on authority investment, over periods of play. Across all periods, average contributions are higher when the authority takes the legitimating action of the big investment.¹⁰ More formally, partitioning the data into the first ten or last ten periods and re-estimating the specification shown in column (2) yields an estimated legitimacy effect of .097 in periods 1-10 ($p = 0.01$, two-tailed) and .130 in periods 11-20 ($p = 0.06$, two-tailed). This evidence does not prove that conditional cooperation is unimportant. However, the relative stability in the effect of the legitimating action over time is at least strongly suggestive that this mechanism is not the primary driving force behind our key result from Study 1.

4 Study 2: Testing the Direct Reciprocity Mechanism

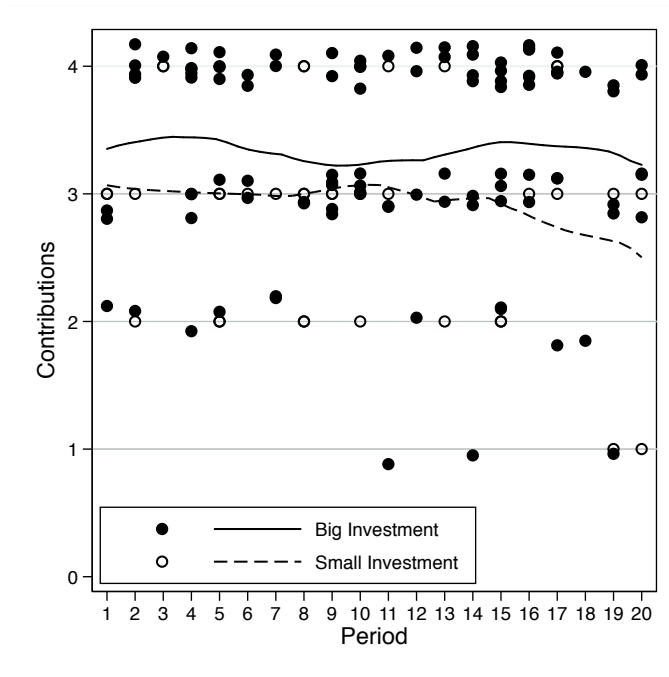
4.1 Design

Our second study is designed to assess whether the citizen to authority reciprocity mechanism described above explains the observed effect of the legitimating action via the intrinsic motivations channel. Our strategy for inquiry into this mechanisms is to slightly vary the setup from Study 1 in a way that makes the presence or absence of differences in citizen behavior informative about underlying mechanisms.

In Study 2, we assess the plausibility of the reciprocity mechanism by breaking the deterministic linkage between the citizen's contribution decision and the authority's material

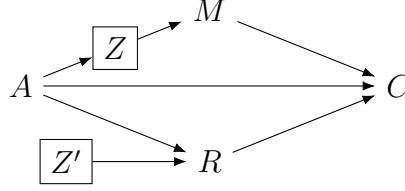
¹⁰The persistence of this effect over time is also reassuring because it mitigates against reciprocity-based explanations. If reciprocity alone explained the pattern we observed, then we would expect it to be harder to sustain support for a big investment in later periods.

Figure 4: Effect of the Legitimizing Action: Persistence Over Time



Note: Data jittered to enhance clarity of presentation. Observations are group contributions given medium accuracy and PATS enforcement strategy. Lines are local polynomial smoothers.

Figure 5: Directed Acyclic Graph Depicting Study 2’s Design



Study 2 introduces a second randomization, labeled Z' , which affects whether the authority materially benefits from the citizens’ contributions to the public good.

welfare. Specifically, we adapt the setup of Study 1 by introducing an additional randomization, which takes place after the authority chooses an investment level and enforcement rule. With probability 0.5, the authority receives a flat fee of 20 tokens but does not benefit from contributions to the common pot; and with probability 0.5, she receives the 20 tokens plus 0.4 times the contributions to the common pot. Figure 5 depicts the underlying logic of the design: here, R denotes reciprocity motivations, and Z' the coin flip that determines whether the authority benefits from citizen contributions.

The randomization denoted Z' takes place after the authority has made her decisions for the period but before the citizens have made their contribution decisions. Therefore, at the time the citizens are deciding whether to contribute to the public good, they both know all material factors that should affect their contribution decisions (as in the original experiment), as well as whether the authority’s welfare will be affected by their contributions. When the authority directly benefits from contributions to the common pot, contributions are compatible with the reciprocity mechanism. By contrast, when the authority is paid only the flat fee, citizens know their contributions have no effect on the authority’s welfare, and thus cannot be affected by these reciprocity concerns.

Study 2 differs from Study 1 in one further respect. We improve the statistical power of the experiment by including only two (rather than three) levels of accuracy: High (20% error rate) and Low (40% error rate). In Study 2, conditional on the small investment, accuracy was Low with 75% probability and High with 25% probability; conditional on the

big investment, those probabilities were reversed. This setup allows us to estimate the effects of the legitimating action, for each compensation mechanism, both under Low accuracy ($\overline{C}_{B,L}^m - \overline{C}_{S,L}^m$), modifying the notation from above with the superscript $m \in \{flat, benefit\}$) and High accuracy ($\overline{C}_{B,H}^m - \overline{C}_{S,H}^m$).

Note that in Study 2, unlike in Study 1, there is no realized level of accuracy that would make a subject motivated purely by extrinsic incentives completely indifferent with respect to contribution. Such a subject would strictly prefer to contribute given high accuracy, and strictly prefer not to given low accuracy. Given prior experimental results suggesting that subjects contribute more in public goods games than would be anticipated based on material incentives alone, ex ante we would expect greater variation in the contribution choice under low than high accuracy: under low accuracy, idiosyncratic motivations to contribute would push subjects closer to indifference, and under high accuracy, further from that indifference.

4.2 Study 2: Experimental Results

We have data from 110 subjects gathered during 7 sessions conducted at [Redacted]. Subjects earned an average of \$19.27 (s.d. of \$2.5), with a maximum of \$25.67 and a minimum of \$12.47. The average score on the quiz administered between the instructions and the experiment was 6.15 out of 8 (s.d. 1.5), with 53% receiving 7 or higher and 17.3% receiving a perfect score.

Aggregate Authority and Citizen Behavior. Table 4 summarizes authority choices across all 440 group-period interactions. As in the first experiment, a majority of enforcers chose the PATS enforcement rule, although by a smaller margin (57% vs. 74%). This may reflect less concern for the welfare of the citizens in one’s group given the lower expected stakes of the choice in this setting. Additionally, fewer subjects in the role of the authority chose the big investment: 33% vs. 60%. Conditional on choosing the PATS enforcement rule, authorities chose the big investment 43% of the time (compared to 69% of the time in the baseline experiment). Overall, citizens contributed to the public good 48% of the time. This rate, while lower than in the first experiment, is similarly stable over time (See Figure

Table 4: Authority Choices in the Second Experiment: Group-Level Data

	Enforcement Rule				Total
	PATS	Anti-PATS	Never Punish	Always Punish	
Small Investment	142	35	82	35	294
Big Investment	108	5	13	20	146
Total	250	40	95	55	440

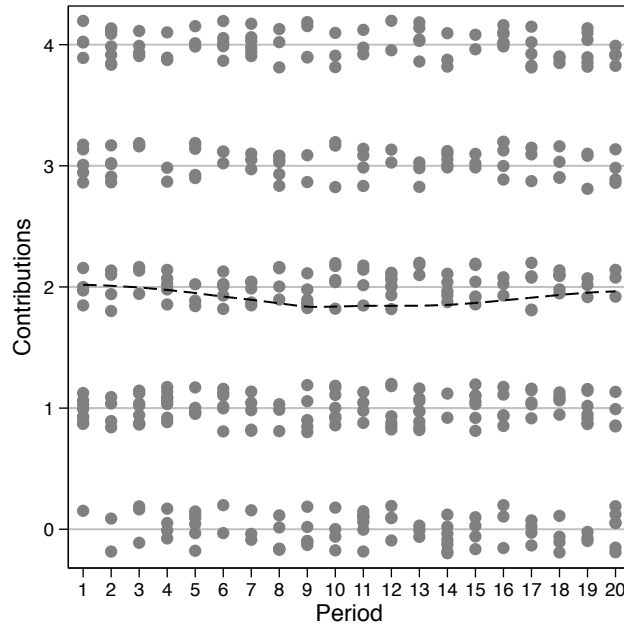
6).

Aggregated Effects. Before proceeding to our main analysis (disaggregating citizen behavior by the authority’s compensation mechanism), we present summary data suggesting that the findings of the first experiment replicate in this alternative environment. Figure 7 displays average group contribution rates conditional on the PATS enforcement rule, for different authority investments and realized levels of accuracy.

The institutional effects are obtained by comparing, respectively, the dark gray (bar #3) and white (bar #1) bars (conditioning on small investment) and the black (bar #4) and pale gray (bar #2) bars (conditioning on big investment). They are unambiguous, and overwhelmingly statistically significant: conditional on the small investment, an increase from low to high accuracy nearly doubles the contribution rate, from 1.74 to 3.35. The institutional effect conditional on the big investment is smaller in magnitude, but still highly significant: an increase from 2.20 to 3.47, or 58%.

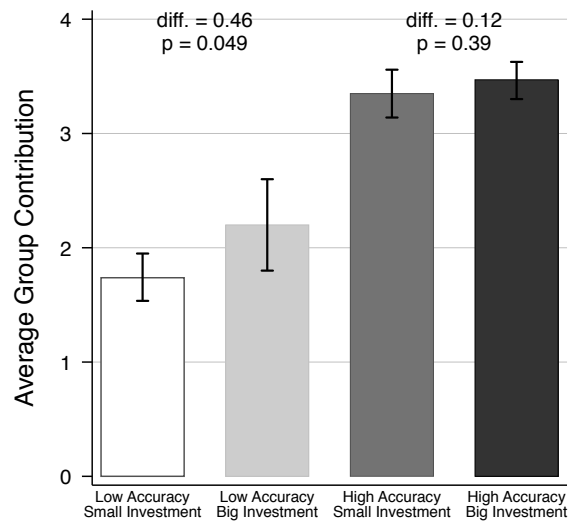
Next, we estimate the effect of the legitimating action, calculated at the group-period level and aggregating across authority compensation mechanisms. The effect conditional on high accuracy is obtained by comparing the black and dark gray bars (bar #4 and bar #3, respectively) in Figure 7. We find the effect to be a statistically insignificant different of 0.12 contributions ($p = 0.39$, two-tailed). Recall that conditional on the high level of accuracy, subjects should strictly prefer to contribute to the public good based on extrinsic motivations alone, so the fact that average contribution rates are high under both levels of authority investment is unsurprising.

Figure 6: Group Contributions by Period, Study 2



Note: Data jittered to enhance clarity of presentation. Dashed line is local polynomial smoother.

Figure 7: Group-level Contributions by Investment Decision and Accuracy Levels Given the Authority Choses Punish According to Signal, Second Experiment



Note: Averages with bootstrapped 95% confidence intervals.

More relevant for the current discussion is the effect of authority investment conditional on the low accuracy level (comparing bar #2 and bar #1), recalling that extrinsic motives alone are insufficient to motivate contributions given this realization. Here, we observe a statistically significant 27% increase in contributions, from an average of 1.74 to 2.2.

Effects of the Legitimizing Action and Authority Compensation. As above, our main analysis focuses on the effect on an individual citizen’s contribution decision of the authority’s investment choice in cases in which the authority chooses the PATS enforcement rule, holding fixed the accuracy level. The added wrinkle is to condition this effect on the authority’s compensation mechanism – which, recall, is realized after the authority’s choices, but before the citizens’. Our analysis appears in Table 5.

Column (1) estimates the effect of the authority’s investment decision by pooling across authority compensation mechanisms and accuracy levels. The specification suggests that the authority’s choosing the big investment is associated with a statistically significant 6.6 percentage point increase in the likelihood of contribution. Confirming the results from the group-level analysis, the effect of accuracy is considerably larger, however, increasing contributions by 37 percentage points (relative to a baseline of 44%). Column (2) adds period-specific effects as well as two additional control variables capturing the subject’s experiences in previous periods.

Column (3) disaggregates the effect of the legitimating action by the authority’s compensation mechanism. If the reciprocity mechanism explains the earlier result, we would expect the coefficient on the interaction between the authority benefiting from the public good and the big investment to be positive while the coefficient on the authority’s investment alone would diminish in size. But this is not what we find. The estimates suggest that when the authority does *not* benefit from the public good, her choice of the big investment leads to a 7.6 percentage point increase in contributions ($p = 0.063$, two-tailed). When the authority does benefit, that rate decreases to 5.8 percentage points and is no longer statistically significant at conventional levels ($p = 0.19$). Column (4) adds the same vector of controls

Table 5: Conditional Effect of Authority's Investment Choice under Different Authority Compensation Mechanisms, Given PATS Enforcement Rule

	(1)	(2)	(3)	(4)
Accuracy (1=high, 0=low)	0.370 [0.032]***	0.403 [0.027]***	0.369 [0.032]***	0.403 [0.027]***
Authority Investment (1=high, 0=low)	0.066 [0.031]**	0.063 [0.028]**	0.076 [0.041]*	0.058 [0.035]
Average group contributions in prior periods (0-4)		0.154 [0.023]***		0.154 [0.023]***
Average investments in prior periods (0-1)		0.001 [0.078]		0.001 [0.078]
Authority benefits authority benefits \times investment			-0.017 [0.042]	-0.026 [0.036]
			-0.019 [0.058]	0.013 [0.050]
Constant	0.444 [0.025]***	0.102 [0.072]	0.452 [0.031]***	0.113 [0.074]
Observations	1,000	960	1,000	960
R-squared	0.181	0.245	0.182	0.246
Period fixed effects	N	Y	N	Y

Dependent variable is player contribution decision (1=yes, 0=no).

OLS coefficients with group-period clustered standard errors in brackets.

Observations are individual contribution decisions given PATS enforcement strategy.

as in the column (2) specification, yielding an estimated effect of the authority’s investment of 5.8 percentage points ($p = 0.10$) when the authority does not enjoy the public good, and a slightly larger effect of 7.1 percentage points when she does benefit ($p = 0.073$). In neither specification (3) nor (4) are the estimated effects of the authority’s investment choice statistically distinguishable from one another.

5 Discussion

This paper makes several contributions to the study of legitimate authority and its relationship with subordinate behavior. Foremost, we demonstrate that subordinates can be motivated to comply with an authority as a consequence of changes to her perceived legitimacy, holding constant purely instrumental motivations. This phenomenon contributes fundamentally to our understanding of how perceptions of institutional fairness affect behavior, a dynamic that persistently plays out in salient policy domains such as policing.

Methodologically, the contribution of this paper is to formalize the challenge of separately identifying legitimating effects from material motivations for compliance, and to propose an experimental protocol that solves this attribution problem. The innovation of our experimental designs is to isolate the legitimating effect of an authority’s effort to secure a more procedurally fair institution from the effects of the institution itself. We do so by making the legitimating *action* of the authority probabilistically, rather than deterministically, related to the materially- and institutionally-relevant *consequences* of that action.

In our experiment, we find that an authority’s legitimating action substantially increases the willingness of citizens to contribute to a collective good. This effect is robust to different sample restrictions and statistical modeling approaches. In a follow-on experiment we show that this effect cannot be explained by an analytically distinct mechanism in which citizens “repay” an authority for her efforts to improve citizen welfare.

This study departs from the corpus of prior research on procedural fairness and other sources of legitimacy in its use of an incentivized laboratory environment in which subjects receive financial compensation for their performance in a game. This approach offers several

important advantages. First, it permits us to establish a benchmark of rational behavior based on purely material motivations against which to compare actual behavior. With this benchmark in hand, we can more definitively attribute observed differences in contribution behavior to specific psychological motivations. Second, because subjects benefit or suffer materially from their actions and those of other subjects, our experimental environment approximates the sorts of compliance choices that individuals must make in their day-to-day lives.

Third, our specific experimental design also allows us to make progress in understanding the psychological origins of legitimacy. We show that authorities who attempt to obtain more accurate institutions enjoy more legitimacy with their citizens. Of course there are numerous potential sources of improved legitimacy aside from accuracy. For example, the authority may or may not take actions that appear biased against a member of a specific group. The experimental approach presented here is sufficiently flexible to study other sources of legitimacy, a task for future research.

The stylized environment in which we are able to implement this design may depart from non-experimental settings in a number of respects. Two features of the authority essential to the design that deserve comment are her ability to pre-commit to an enforcement rule and her unlimited enforcement capacity. In the absence of pre-commitment, citizens would be able to make inferences about likely enforcement strategies from the (ex ante) procedural investment, further entangling instrumental and legitimacy-based mechanisms. Likewise, if an authority's capacity is limited, in many circumstances there will be multiple equilibria in which expectations about others' behavior come into play. In such an environment, authorities' legitimating actions could serve an equilibrium selection role, affecting citizen behavior by coordinating expectations about others' behavior.

We conclude by noting the broader value of isolating different causal mechanisms associated with policies aimed at fostering cooperation and compliance. One might be tempted to argue that if some institution "works," it is irrelevant whether it does so owing to its

intrinsic or extrinsic effects. But particularly given the frequency with which legitimacy is invoked causally, this risks serious misunderstanding. This misunderstanding may itself be undesirable, but if it in turn leads to bad policy advice, the consequences may also be deleterious from a public welfare perspective. In particular, there are many situations in which policymakers confront choices about which (costly) reform to implement. While it is often the case that enhanced legitimacy and improved material motivations for compliance move hand in hand, they do not always do so (e.g., in the case of policies surrounding racial profiling or other forms of group targeting) and sometimes interventions that affect one causal pathway more than another are differentially costly. Fully isolating and understanding the empirical consequences of legitimacy for compliance allows us to better make predictions and recommendations in situations in which the contours of policy involve choices along these lines.

References

- Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics* 97 (4): 543–569.
- Anscombe, G.E.M. 1981. *The Collected Philosophical Papers of G.E.M. Anscombe*. Oxford: Basil Blackwell.
- Baldassarri, Delia, and Guy Grossman. 2011. "Centralized sanctioning and legitimate authority promote cooperation in humans." *Proceedings of the National Academy of Sciences* 108 (27): 11023–11027.
- Bottoms, Anthony, and Justice Tankebe. 2012. "Beyond Procedural Justice." *Journal of Criminal Law and Criminology* 102 (1): 119–170.
- Caldeira, Gregory A., and James L. Gibson. 1992. "The Etiology of Public Support for the Supreme Court." *American Journal of Political Science* 36 (3): 635–664.

- Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior." *American Economic Review* 100 (5): 2205–29.
- Department of Justice. 2015. "Investigation of the Ferguson Police Department."
- Easton, David. 1965. *A Systems Analysis of Political Life*. New York: Wiley.
- Fehr, Ernst, and Simon Gächter. 2000. "Cooperation and Punishment in Public Goods Experiments." *The American Economic Review* 90 (4): 980–994.
- Fischbacher, Urs. 2007. "z-Tree." *Experimental Economics* 10 (2): 171–178.
- Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are people conditionally cooperative?" *Economics Letters* 71 (3): 397 – 404.
- Gibson, James L., Gregory A. Caldeira, and Vanessa A. Baird. 1998. "On the Legitimacy of National High Courts." *The American Political Science Review* 92 (2): 343–358.
- Grossman, Guy, and Delia Baldassarri. 2012. "The Impact of Elections on Cooperation." *American Journal of Political Science* 56 (4): 964–985.
- Habermas, Jürgen. 1979. *Communication and the Evolution of Society*. Boston, MA: Beacon Press.
- Hurd, Ian. 1999. "Legitimacy and Authority in International Politics." *International Organization* 53 (2): 379–408.
- Hyde, Alan. 1983. "The Concept of Legitimation in the Sociology of Law." *Wisconsin Law Review* 1983: 379–426.
- Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. "Experimental designs for identifying causal mechanisms." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (1): 5–51.

- Ledyard, John. 1995. "Public Goods." In *Handbook of Experimental Economics*, ed. John H. Kagel, and Alvin E. Roth. Princeton, NJ: Princeton University Press pp. 111–194.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley, CA: University of California Press.
- Lind, E. Allan, and Tom R. Tyler. 1988. *The Social Psychology of Procedural Justice*. New York: Plenum Press.
- Murphy, Kristina. 2005. "Regulating More Effectively." *Journal of Law and Society* 32 (4): 562–589.
- Myerson, Roger B. 2004. "Justice, Institutions, and Multiple Equilibria." *Chicago Journal of International Law* 5 (1): 91–107.
- North, Douglas C. 1981. *Structure and Change in Economic History*. New York: W. W. Norton & Company.
- Paternoster, Raymond, Robert Brame, Ronet Bachman, and Lawrence W. Sherman. 1997. "Do Fair Procedures Matter?" *Law and Society Review* 31 (1): 163–204.
- Pew Research Center. 2016. "On Views of Race and Inequality, Blacks and Whites Are Worlds Apart."
- Pirlott, Angela G., and David P. MacKinnon. 2016. "Design approaches to experimental mediation." *Journal of Experimental Social Psychology* 66: 29 – 38.
- Raz, Joseph. 2009. *The Authority of Law*. 2nd ed. Oxford: Oxford University Press.
- Tyler, Tom R. 1990. *Why People Obey the Law*. New Haven, CT: Yale University Press.
- Tyler, Tom R. 1997. "The Psychology of Legitimacy." *Personality and Social Psychology Review* 1 (4): 323–345.
- Tyler, Tom R., and Yuen J. Ho. 2002. *Trust in the Law*. New York: Russell Sage Foundation.

- van den Bos, Kees. 2001. "Uncertainty Management." *Journal of personality and social psychology* 80 (6): 931–941.
- Vermunt, Riël, Arjaan Wit, Kees van den Bos, and E. Allan Lind. 1996. "The effects of unfair procedure on negative affect and protest." *Social Justice Research* 9 (2): 109–119.
- Weber, Max. 1978. *Economy and Society*. Berkeley, CA: University of California Press.