"Just Words? Just Speeches?"
On the Economic Value of Charismatic Leadership

John Antonakis (University of Lausanne)

Giovanna d'Adda (Università degli Studi di Milano)

Roberto A. Weber (University of Zurich)

Christian Zehnder (University of Lausanne)

June 2021

Abstract

Leadership theories in sociology and psychology argue that effective leaders influence follower behavior not only through the design of incentives and institutions, but also through personal abilities to persuade and motivate. Although charismatic leadership has received considerable attention in the management literature, existing research has not yet established causal evidence for an effect of leader charisma on follower performance in incentivized and economically relevant situations. We report evidence from field and laboratory experiments that investigate whether a leader's charisma—in the form of a stylistically different motivational speech—can induce individuals to undertake personally costly but socially beneficial actions. In the field experiment, we find that workers who are given a charismatic speech increase their output by about 17% relative to workers who listen to a standard speech. This effect is statistically significant and comparable in size to the positive effect of high-powered financial incentives. We then investigate the effect of charisma in a series of laboratory experiments in which subjects are exposed to motivational speeches before playing a repeated public goods game. Our results reveal that a higher number of charismatic elements in the speech can increase public-good contributions by up to 19%. However, we also find that the effectiveness of charisma varies and appears to depend on the social context in which the speech is delivered.

1 Introduction

"Just words, just speeches" was a refrain used in a speech by former U.S. President Barack Obama, and earlier by former governor Deval Patrick. They used it to highlight the importance of leader rhetoric in giving hope, direction, and meaning to followers. Such soft means of leader influence have long held allure as a potentially powerful influence on human behavior and featured prominently in sociological and social psychological theories of organizations and bureaucracies (Barnard, 1940; Katz & Kahn, 1978; Weber, 1947). Leadership research often parallels these perceptions, noting that leaders can influence follower behavior not only through the design of incentives and institutions, but also through personal abilities to persuade and motivate (House, 1977). Hence, a leader's "charisma"—the ability to transmit information in a symbolic, value-based, and emotional manner (Antonakis et al., 2016)—is a potentially important channel through which leaders exert influence.

However, the question of whether more charismatic "words" can truly motivate and inspire followers remains in need of greater scrutiny. Although the fields of psychology and management have contributed much to understanding leadership, there are still many limitations that bound what conclusions we can draw from the existing empirical work in these fields. For example, field studies predicting outcomes from measured leadership styles typically ignore that the measures are endogenous and that their correlations with outcomes are very plausibly due to omitted variables (see Antonakis et al., 2010, for a detailed discussion). The "identification revolution" from economics (cf. Angrist & Pischke, 2010) has not yet made its mark in the management and leadership literatures, despite a growing group of scholars increasingly noting that causal identification is often ignored and that policy implications from endogeneity-plagued findings are few (Gottfredson et al, 2020; Güntner et al., 2020; Sajons, 2020; Sieweke & Santoni, 2020).

Moreover, whereas psychologists have studied manipulations of charisma in the laboratory (e.g., Howell & Frost, 1989; Shea & Howell, 1999; Towler, 2003), the tasks on which participants have worked tend to be trivial, with hypothetical or low stakes, and are never compared to incentivized counterfactual conditions. In addition, there is a paucity of robust evidence on the effects of charismatic leadership from field experiments (Eden, 2020). Thus, despite thousands of published articles in the leadership field, causal evidence for an effect of leadership in general—and leader charisma in particular—on follower performance in incentivized and economically relevant situations has yet to be convincingly demonstrated.

In parallel to research in leadership and management, economists have primarily emphasized incentives and contract design as the primary mechanism through which principals influence agents (Lazear & Rosen, 1981; Holmström, 1982; Holmström & Milgrom, 1991; Lazear, 2000); this mode of influence is usually referred to as "transactional leadership" in the political science, management and psychology literatures (Bass, 1985; Burns, 1978). Economic research has, for the most part, largely ignored the "softer" channels, including charismatic communication, through which other social scientists assume an important part of leader influence occurs. The limited study of leadership mechanisms in economics primarily focuses on situations where a party with private information takes costly actions to convey information and thereby influence behavior (Hermalin, 1998; Potters et al., 2007). Within the standard economic framework, it is difficult to model how words alone—without conveying costly signals—can motivate people to change their behavior. Indeed, a recent theoretical model of how "charismatic" leaders influence followers (Hermalin, 2017) assumes that at least part of this influence is due to the possibility of irrational responses to charisma.

A few experimental studies in economics provide causal evidence on the extent to which "just words" directly influence followers' actions.[1] Most relevant, for our purposes, is work by Kvaløy et al. (2015), who use a field experiment to test whether the inclusion of written motivational sentences increases performance in a data entry job, finding that such statements have a positive effect only when combined with financial incentives. Related work on persuasion shows experimentally how interacting with political leaders affects citizens' political preferences and behaviors (Minozzi et al. 2015). Similarly, "moral suasion"— messages espousing moral arguments for contributing to a public good—can increase voluntary contributions, although with effects that diminish over time (Bott et al. 2020; Dal Bo and Dal Bo 2014). Finally, d'Adda, et al (2017) find that letting leaders communicate with followers often leads to more cheating and dishonesty. Hence, there is some evidence that words alone can have an impact on economic behavior, but this evidence is limited and mixed.[2]

---

[1] A recent paper by Bénabou et al. (2018) investigates theoretically how statements—in the form of "narratives" or "imperatives"—can influence moral behavior. Indeed, as in the psychological literature on leadership, they allow the possibility that the style and rhetoric through which a message is delivered can lead people to adopt different beliefs about the degree to which a costly act is moral.

[2] Fest et al. (in press) use a similar design to that in Kvaløy et al. (2015) to explore the effects of written messages that either convey positive expectations or set specific goals. Neither type of message has a positive effect on performance in their study. A handful of laboratory studies provide correlational evidence that certain types of messages from leaders, making substantively different kinds of arguments, can be more effective for inducing change to more efficient equilibria in coordination games (Brandts & Cooper, 2007; Brandts et al., 2015). However, the investigation of message content in this research is post hoc and provides no evidence on variation in style or charisma of leaders' messages.

Our paper attempts to bridge the gap between economics and sociological and psychological research on charismatic leadership. Our research therefore has a different focus than the previous literature: all above-mentioned studies in economics either compared the effect of having a motivating message to not having such a message or contrasted motivating messages with varying substantive content. Our studies, in contrast, explore how the delivery and style of leaders' communication—and particularly varying the charisma of such communication—shapes the effect of a particular motivating message on performance. In other words, we are interested in testing whether a leader who delivers a message "better"—that is, in a more charismatic way—has a more powerful influence on followers' efforts.

We first present a field experiment intended to test, whether "charismatic" leadership, in the form of a stylistically different speech given by a leader, can induce workers to exert higher effort. In a second step, we further explore the generalizability and limitations of the motivating effect of charisma in a series of laboratory experiments in which participants are exposed to varying speeches before participating in a repeated public goods game.[3] For a clean identification of the charisma effect, we always let the same individual (a trained actor) deliver the two speeches and we ensured that the number of words and the content—that is, the de facto topics, information, structure and themes presented—of the paired speeches were qualitatively very similar. Our objective was to vary only the message delivery, in the form of the leader's use of charismatic communication techniques identified in previous psychological leadership research. To ensure that the speeches were perceived as intended, we performed manipulation checks in which independent samples of participants evaluated each speech along several dimensions.

In our field experiment, temporary workers have to prepare envelopes for a fundraising campaign conducted on behalf of a hospital. Workers are exposed to short motivational speeches that differ in the number of charismatic elements. To quantify the impact of charisma on worker performance, we also compare its effectiveness with the effect of high-powered performance pay. As expected, we find that performance pay significantly increases workers' performance, by about 20%, compared to a baseline condition, an increase that is similar in size to those found in other field studies (e.g., Lazear, 2000; Shearer, 2004). More interestingly, our charismatic motivational speech increases performance by about 17% relative to the baseline low-charisma condition, a statistically significant effect that is indistinguishable from that of financial incentives. Moreover, whereas the piece rate condition engenders a small and

---

[3] Our work builds on a large body of interdisciplinary studies that introduce insights from diverse areas of social science to understand the factors that influence human cooperation (e.g., Deutsch, et al., 1967; Dawes, et al. 1977).

insignificant increase in the cost per letter, the charisma condition significantly decreases the cost per letter by about 15%. This calculation of the reduced cost per letter does not take into account that the provision of a charismatic speech might be costly as well, but our data allow us to determine the maximal amount that could be spent so that charisma remains cost effective. Quality checks suggest that the positive effects of charisma and financial incentives do not come at the cost of lower quality.

The field experiment provides clear evidence that varying charisma *can* influence the extent to which leader communication induces desirable follower actions. However, to develop a robust understanding of the effects of charisma it is important to test its influence in other settings and to develop paradigms that allow researchers to more easily study its effectiveness in contexts with varying features. We thus complement our field experiment with a series of laboratory studies. The laboratory experiments establish an easily replicable and flexible workhorse setting that we and others can use for future research on charisma.

In all our laboratory experiments, participants play the same four-person public goods game for 10 periods. There are many important differences between this laboratory setting and the context we use to study the effects of leader charisma in the field. Our objective was not to try to identify a laboratory setting as similar as possible to the field context. Rather, we aimed to study the effects of charisma on a behavior—voluntary cooperation—that is of fundamental importance in the social and organizational sciences using a laboratory paradigm that has been previously studied extensively. Our philosophy was thus to start with a very stark environment that mimics the way public goods experiments are typically conducted and that omits many features of our field experiment. Doing so allows us to identify boundary conditions, beyond which charisma is unlikely to have an effect—and where the observation of an effect would be surprising. We then add features that are present in the field experiment to shed light on factors that might play a role in determining whether charisma has a motivating influence.

In all laboratory experiments, we compare a high charisma condition, where participants watch a charismatic speech before making the first contribution decision, to a low charisma condition, where participants see a similar speech, but presented in a less charismatic manner. In the first laboratory experiment, we add a baseline condition in which participants do not observe a speech. In the second laboratory experiment, we increase the moral dimension of the contribution decision—thereby bringing it closer to the context in our field experiment—by modifying the game so that each contribution to the public good also creates a donation to a children's hospital. In the third laboratory experiment, participants watch the video as a group rather than individually, reproducing the social context from our field experiment.

We find that the high charisma condition yields at least slightly higher contributions than the low charisma condition in all laboratory experiments. However, the effect size and its statistical strength vary considerably across studies. The biggest difference arises when followers are exposed to a charismatic leader in each other's presence, a feature that is also present in our field experiment.[4] Contributions to the public good in the third laboratory experiment are about 19% higher in the charisma condition compared to no charisma—an effect very similar to that in our field experiment—and this difference is statistically significant. We use the observation that the effects of charisma appear strongest in our study when individuals observe one another receiving the leader's speech, along with other aspects of our data, to cautiously speculate about the mechanisms that may influence the effects of charisma on follower behavior. For example, we use measures of beliefs of others' contributions from our laboratory experiment to rule out a direct influence of charisma on beliefs as the main channel. Instead, dynamic aspects of the data suggest that groups exposed jointly to charismatic communication exhibit greater resilience in not decreasing contributions over time.

To the best of our knowledge, our experiment is the first to document a causal effect of leader charisma on follower behavior in economically relevant environments. Previous studies have established that non-monetary motivational instruments, such as symbolic awards (Kosfeld & Neckermann, 2011; Bradler et al., 2016) or messages about the meaning of work (Grant, 2008; Kosfeld et al., 2014) can increase workers' output. There is also evidence that nudging leadership in non-routine team-tasks substantively increases performance (Englmaier et al., 2019). As we note earlier, other studies find some positive impact of brief motivational statements (Kvaløy et al., 2015; Dal Bó & Dal Bó, 2014). However, we show that it may not only be important to provide followers with such statements, but that how the message is delivered can also have significant impacts. If leaders use charismatic leadership tactics, the same substantive message can have a much larger positive impact on followers' behavior.

The remainder of the paper is structured as follows: In the following section, we discuss the concept of charisma in more detail. In this section we also present an ad hoc model of how a leader's charisma might influence follower behavior, which we return to in developing concrete hypotheses for each experiment. Section 3 presents the design and results for our field experiment, while Section 4 presents the laboratory experiments. In Section 5, we discuss the combined evidence from our studies and provide a broad conclusion.

---

[4] This finding is partly consistent with the model of Hermalin (2017), where part of the effectiveness of charisma comes from a rational individual's belief that other irrational followers will be influenced by charisma.

2 Charisma and Leadership

2.1 Background

Because the concept of charisma is rarely used in economic research, we find it useful to provide some background (see Antonakis et al., 2016 for a detailed historical overview). The term "charisma" (from the Greek "χάρισμα," meaning gift) was first used by Max Weber (1947) to describe leaders who appeared to have "supernatural, superhuman, or at least specifically exceptional powers or qualities" (p. 358). Most of the early scholarship on charisma, descriptive in nature, was focused in the political science literature (Burns 1978; Downton 1973; Willner 1984). House (1977) first proposed a detailed psychological account of how charismatic leaders impact followers, rejecting the notion that charisma was somehow a mysterious quality. For House, it was important to understand the basis of the psychological interaction that charismatic leaders have with their followers, resulting from the characteristics of the leader (e.g., confidence and moral conviction), what the leader says, and the situation in which the influencing process occurs.

The first systematic empirical research program on charisma began in the field of applied psychology, led by Bernard M. Bass (e.g., Bass, 1985; Bass et al., 1987; Waldman et al., 1990); other research groups followed suit (e.g., Alimo-Metcalfe & Alban-Metcalfe, 2001; Conger et al., 1997; De Hoogh et al., 2004; Podsakoff et al., 1990). The psychological approach has generally used questionnaires to gauge perceptions of charisma (see Antonakis et al., 2016 for details), with the best known one referred to as the Multifactor Leadership Questionnaire (MLQ) (Bass & Avolio, 1995). This research strand follows from the idea that charisma is in the eye of the beholder and thus measuring perceptions of charisma in observers captures true variation in the behaviors signaled by the leader.

This approach to studying charisma has been strongly challenged, initially by Yukl (1999), and later by Antonakis et al, (2011, 2016) and van Knippenberg and Sitkin (2013). These critiques identified limitations in the study of charisma including that: 1) charisma was confounded with a broader leadership typology called transformational leadership, which includes other fundamentally different leadership approaches (e.g., being individually considerate) that have nothing to do with charisma per se, 2) charisma was defined and measured in terms of its outcomes and was thus conceptually and empirically confounded and 3) questionnaire measures of charisma are endogenous—they are biased by omitted variables; these variables include a host of causes at various levels of analysis (e.g., leader, observer, context, microeconomic environment) that may also affect the measure outcome. While it may

seem clear that such measures cannot be modeled as causes of other variables, the ease with which questionnaire data can be gathered makes it an appealing approach in much organizational research. However, the obvious limitations in modeling these measures as ostensive causes of consequential outcomes has been noted in the management literature (Fischer et al., 2020).

The effects of charisma have been isolated and exogenously manipulated in the laboratory, producing valuable contributions on the effects of charismatic leadership (e.g., Howell & Frost, 1989; Shea & Howell, 1999; Antonakis et al., 2011; Frese et al., 2003; Towler, 2003). Yet, as mentioned earlier, these experiments lacked consequential outcomes and the treatment effects have never been compared to strong counterfactual conditions (e.g., against economic incentives). Still, this literature as well as other work (e.g., Den Hartog & Verburg, 1997; Shamir et al., 1994) provides a useful starting point for identifying communication techniques of charismatic leaders.

Turning to the precise characteristics that make a leader charismatic, Antonakis and colleagues (Antonakis et al., 2011; Jacquart & Antonakis, 2015; Tur et al., 2021) identify nine verbal and three non-verbal "charismatic leadership tactics" (CLTs). The CLTs are thought to help make a leader's message more salient by making it easier to picture and remember, and by arousing follower emotions. From an experimentalist's point of view, the CLTs are interesting because they can be manipulated. The tactics include: (i) metaphors, to simplify the message and trigger easy recall; (ii) rhetorical questions, to create intrigue and suspense and direct attention to seeking the answer; (iii) stories and anecdotes, to trigger imagery and recall, and identify a relevant moral; (iv) contrasts, to define what should be done versus what should not be done and (v) three-part lists, to provide sufficient proof or completeness; (vi) expressing moral conviction, to focus attention on moral justification and on doing what is right; (vii) expressing the sentiments of the collective, essentially using cognitive empathy to engender identification (via similarity) with the leader; (viii) setting high and ambitious goals, to make followers feel competent and focus their effort on a target and (ix) creating confidence that goals can be achieved. Charismatic leaders also affect followers via non-verbal means that reinforce the message and convey the leaders' emotional states, including (x) body gestures, (xi) facial expressions, and (xii) an animated voice tone (for more details on the CLTs and examples see Antonakis et al. 2012). In naturally-occurring data the presence of verbal and non-verbal CLTs in leaders' communication tend to correlate quite strongly (r = .48, Antonakis et al. 2011). We will build on this research stream to robustly test the impacts of charisma.

2.2 Effects of Leader Charisma on Follower Behavior: An Ad-Hoc Model

In standard economic theory, stylistic communication—the way in which the content is delivered—plays no role, because only informational content matters. We present a simple model that illustrates the potential impact of leader charisma on follower performance. The model does not capture the details of the complex psychological mechanisms underlying leader-follower interactions, but rather relies on an *ad hoc* framework. The model is useful because it formalizes our intuitions and delivers testable predictions.

We model an environment in which N followers choose an action in response to a motivational message of a leader. Follower i's utility function has the following form:

$$U_i = m(\boldsymbol{a}) + n(\boldsymbol{a},c) - k(a_i),$$

where a_i is follower i's productive action, $\boldsymbol{a}$ is a vector containing all followers' actions ($\boldsymbol{a} = \{a_1, a_2, \ldots, a_N\}$), c is a measure of the leader's charisma as perceived by the followers, m is the follower's material benefit, n is the intrinsic (or non-material) utility that the worker derives from performing action a_i, and k is the follower's effort cost.

We assume that the follower's material benefit m weakly increases in the follower's own action ($\partial m/\partial a_i \geq 0$), and in other followers' actions ($\partial m/\partial a_j \geq 0$). The follower's intrinsic utility n can either be a positive term (the follower may enjoy contributing to a useful cause) or a negative term (the follower may feel guilty about not contributing to a good cause).[5] This type of utility may not only be driven by the follower's own action a_i, but may also depend on the contribution of others $\boldsymbol{a}_{-i}$ and the perceived leader charisma c. We assume that the worker's productive action has a strictly positive impact on intrinsic utility ($\partial n/\partial a_i > 0$ and $\partial^2 n/\partial a_i^2 \leq 0$), either because contributing makes the worker feel better (if $n > 0$) or because contributing reduces the feeling of guilt (if $n < 0$). Being exposed to a more charismatic leader is assumed to intensify the psychological impact ($\partial n/\partial c > 0$ if $n > 0$ or $\partial n/\partial c < 0$ if $n < 0$) and to increase the intrinsic motivation to perform the task ($\partial^2 n/\partial a_i \partial c > 0$).[6] Moreover, depending on the situation, we allow followers' actions to be strategic complements ($\partial^2 n/\partial a_i \partial a_j \geq 0$) or strategic substitutes ($\partial^2 n/\partial a_i \partial a_j \leq 0$).[7] The cost function k increases in the follower's action ($\partial k/\partial a_i > 0$).

---

[5] For our predictions it makes no difference whether the intrinsic utility is positive or negative. The impact of charisma on motivation may either be driven by an increase in the utility of contributing or by an increase in the disutility of not contributing. The sign may be important in contexts where participation constraints are relevant, but in all our experiments participants commit to participation before being exposed to the treatment.

[6] See Kvaløy & Schöttner (2015) for a related assumption. In their model the principal's "motivational effort" reduces the worker's marginal effort cost.

[7] Strategic complementarity is a plausible assumption in social dilemma settings where conditional cooperation has been shown to play a crucial role (see, e.g., Fischbacher et al. 2001). However, other settings may be

Our assumption imply that charisma may not only have a direct impact on intrinsic motivation ($\partial^2 n/\partial a_i \partial c > 0$), but may also operate through other channels. If the material or intrinsic benefits also depend on others' actions, charisma may also have indirect effects, because each follower anticipates the impact of charisma on her fellow followers (e.g. if $\partial^2 m/\partial a_i \partial a_j \neq 0$ and/or $\partial^2 n/\partial a_i \partial a_j \neq 0$). We will discuss the relevance of these different channels when we present each of our experiments.

3 Field Experiment

3.1 Recruitment and Procedures

We conducted our field experiment in collaboration with the Birmingham Children's Hospital (BCH). In November 2013 the BCH launched their "Super Santa" fundraising campaign to collect money to help children remaining in the hospital during the Christmas period. Donations would fund the purchase of gifts, the refurbishment of rooms and parents' travel expenses, so that the children could spend Christmas with their families.

To launch the campaign the BCH required about 30,000 stuffed donation letters. To prepare these letters we contracted a large temporary employment agency to hire 120 temporary workers.[8] Following their standard procedure, the agency sent the job advertisement to the usual pool of job searchers and posted it at local job centers. The advertisement mentioned that the job would consist of stuffing envelopes for 3 hours at home, that the material would have to be picked up at the agency's office and that each worker would be compensated for 3 hours of work plus 1.5 hours of traveling and instruction time (see Figure 1 for the original job advertisement). The agency handled the sign-up procedure and assigned time slots so that not more than eight workers at a time would show up to collect their materials. This assignment was mostly followed: only one session had more than 8 participants, while the remaining ones had 8 or fewer participants, depending on how many workers failed to show up to the session. Workers were not aware that the job was also part of a scientific study.[9]

---

characterized by strategic substitutability, as when a fixed overall contribution is needed or when thresholds matter. In such contexts, charisma may have diminished impact.

[8] The size of the sample was limited by logistical and implementation constraints. First, we could only have a limited number of workers at a time to receive instructions in a conference room. Second, to have the same actor deliver the speech to all groups, it was infeasible to have multiple simultaneous groups. Third, because workers returned their boxes the following day, we wanted all workers to complete their first session by the time workers from earlier sessions started returning their materials. Fourth, to prevent interactions of workers between sessions, sessions started once per hour. These constraints limited us to 15 sessions consisting of at most 8 workers and also informed our choice to divide the sample into 3 treatments, rather than the 4 of a 2x2 factorial design.

[9] All information about the job was correct and the work was real (all stuffed letters were sent out to potential donors). In line with the requirements of the IRB of the University of Birmingham, we debriefed our participants after the data collection had been completed. The debriefing letter is displayed in Appendix C.

[Figure 1 about here]

Each worker was required to show up twice at the agency's office in Birmingham. On the first day workers participated (in groups of eight) in an instruction session of 25 minutes. An instructor explained the task in detail and each worker assembled a sample letter. At the end each worker signed up for a time slot to return the stuffed letters on the following day[10] and picked up materials to stuff up to 350 letters.[11] The materials were packed in two cardboard boxes (with a weight of about 3.5 kilograms each) that were placed in plastic bags so that the workers could easily carry them. Before the workers left, they were informed that an organizer would provide them with more information about the importance of their job. They were then given a five-minute motivational speech by a professional actor.

Once the instruction session was over, workers took the material home to complete their work there. On the next day each worker returned the completed envelopes during the agreed upon time slot. A researcher then counted the completed envelopes returned by each worker.

3.2   Work Task

The workers' task consisted of stuffing envelopes, placing them neatly into boxes in bundles of ten envelopes and keeping records of printing mistakes. Stuffing an envelope included the following steps. First, the worker had to examine a letter for printing faults. Faulty letters required removal and recording the type of printing fault on a list. Faultless letters had to be neatly folded and placed in a donation envelope. Next, the worker had to handwrite a code from a list on a donation form that was then placed behind the letter in the donation envelope. Finally, the worker had to insert a smaller return envelope and seal the donation envelope. Once ten envelopes had been completed, the worker had to label and bundle them using a rubber band. The bundles had to be packed back into the boxes, so that the worker could return the completed work in good order. See Appendix A for photos of materials and completed bundles.

The advantage of letting workers complete the task at home is that they were separated from each other so that direct interaction and spillover effects can be excluded. Logistically, it would have been challenging to create such an isolated work environment in a natural manner on site. Obviously, we could not observe how workers completed their work at home. This lack

---

[10] A maximum of four people in total, and not more than one person per session, could sign up for each slot. Once a slot was full, it was taken off the list of available time slots.

[11] We determined the quantity of campaign material to be given to each worker with two main objectives. First, we wanted to ensure that the average worker would not be able to complete all the material in the boxes within the time set for the work; pre-test data, which we used to finalize the quantity of material to give each worker, suggested an average productivity of 217 envelopes in three hours. Second, we did not give more material than what was possible for a person of normal build to easily carry to and from their home.

of control is not a problem for the treatment comparison in which we are interested, because it is non-differential by condition. Our main variable of interest is the number of completed envelopes that the workers returned to the agency's office—irrespective of how much time they spent doing their task or whether they coopted others to help them. However, we also separately measured the quality of the work. In addition, we used the waiting time when workers returned their envelopes for a short questionnaire in which we elicited information on workers' age and gender and asked them how long they worked and whether they had received help. However, as these data are self-reported and cannot be verified, we interpret them with caution.

3.3 Treatments

We now describe the three treatment conditions that we implemented in our experiment:

Baseline. Workers received a fixed wage (GBP 28.40) corresponding to the British minimum hourly wage of GBP 6.31 for 4.5 hours (3 hours of work plus 1.5 hours of instructions and travel time). Workers were given the standard motivational speech, which informed in detail about the purpose of the donation campaign and described the importance of the work.[12]

Charisma. The charisma condition was identical to the baseline, except that we replaced the motivational speech with a substantively similar speech delivered more charismatically. The two speeches have the same number of words, their content is very similar and they have been designed to convey the same amount of information and moral conviction. However, the charisma speech relies heavily on the charismatic leadership tactics (CLTs) that we discussed previously, such as metaphors, stories and anecdotes, contrasts, rhetorical questions.

Piece Rate. The piece rate condition was identical to the baseline, except for the payment scheme. The workers in the piece rate condition also received the fixed amount of GBP 28.40. However, workers who completed more than the threshold of 220 envelopes received—in addition to the fixed wage—a piece rate of GBP 0.12 for each additional envelope completed. The workers were informed about the payment scheme during the instruction session.

We designed the incentive scheme based on data that we collected in a pre-test one month before the actual experiment with a sample of 16 workers. In the pretest, workers in the baseline condition completed 217 envelopes, on average. The British minimum wage legislation made a pure performance pay scheme impossible. We therefore adopted a piece rate with a minimum guarantee (see, e.g., Lazear, 2000, for a very similar incentive scheme). Workers whose performance was below the threshold level of 220 envelopes earned the fixed

---

[12] See http://www.hec.unil.ch/jantonakis/charisma.htm for studio versions of all speeches. Full-text transcripts of the speeches and additional information are provided in Appendix B.

wage, whereas workers who stuffed more than 220 envelopes earned the fixed wage plus a piece rate for each additional envelope beyond the threshold. We pushed the piece rate up to the point at which we expected a further increase to be unprofitable (i.e., just below the estimated average cost of a letter in the pre-tested baseline: GBP 28.40 / 217 = GBP 0.13).

3.4   Procedural Details

We ran the experiment on three consecutive days (November 5-7, 2013). We conducted twelve instruction sessions on the first day and another three in the morning of the second day (five sessions for each treatment). The sessions were organized in five blocks of three. Within each block, treatments were randomly assigned to sessions. This procedure guaranteed that the different treatments were equally distributed over the time of day. In total we invited 120 workers and randomly assigned them to sessions. On the first day 88 of the 96 recruited workers showed up and attended the session in full, while on the second day 18 of the 24 recruited workers did. In total we have 106 observations (35 in the baseline condition, 30 in the piece rate condition, and 41 in the charisma condition).[13] Among workers who showed up, some did not attend the session in full or decided not to perform the task: namely, four subjects (two in the baseline and two in the piece-rate condition) left either during or after the session without picking up the material, and one subject in the baseline returned the material untouched one hour after the session. Inability to return the boxes or other work opportunities were mainly cited as reasons for these refusals. In our preferred specifications, we treat these observations as missing, but our results are robust to treating these workers as completing 0 envelopes.[14]

The average age of participating workers was 32.5 years and 55 percent were female. Table 1 presents summary statistics of workers' characteristics and parametric balance tests: the treatment sub-samples are balanced across all characteristics for which we have data.[15]

Workers who received instructions on day 1 brought back their envelopes in the afternoon of day 2; those who were instructed in the morning of day 2 returned in the morning of day 3. On arrival, a collaborator counted the number of completed envelopes and filled out a payment slip. The material was then counted again and quality-checked; namely, the

---

[13] Our sample size is in line or larger than that of similar studies, such as Kube et al. (2012) and Gneezy and List (2006). When discussing the empirical results, we report Minimum Detectable Effects (MDEs) to discuss ex-post power.

[14] We targeted 40 subjects per condition, to be achieved by running 5 sessions with 8 subjects per session in each condition. In order to be able to run sessions even if not all workers turned up, we over-recruited workers for each session. The number of workers attending a session ranges between a minimum of 5 and a maximum of 11. These variations in attendance by session explain the deviations from the target number of subjects by condition.

[15] Balance is confirmed also if we use non-parametric k-sample tests on the equality of medians. None of the Fisher's exact test statistics reveal statistically significant differences across treatment samples.

experimenter examined whether bundles and boxes were neatly packed, and drew three sample envelopes to check whether they were correctly stuffed. We use these checks to construct an index of quality. The payment of the workers was entirely handled by the employment agency.

3.5   Manipulation checks

We conducted two types of manipulation checks using external evaluators so as to not make the workers aware of the experiment (see Lonati et al., 2018): an objective (counting the presence of CLTs) and two subjective ones (measuring perceived charisma).

For the objective manipulation check, two research assistants (trained in recognizing the verbal and non-verbal CLTs) independently coded each speech for the absolute presence of each tactic (see Appendix D). The presence of the nine verbal CLTs was coded at the sentence level (using the transcripts of the speeches).[16] The number of CLTs used as a proportion of total sentences in the low-charisma speech was 39.02% (i.e., 16 tactics over 41 sentences); in the charismatic speech it was 91.67% (i.e., 44 tactics over 48 sentences). This difference is statistically significant ($z = 5.28$, $p < 0.001$, Koopman, 1984). Although our focus was mostly on verbal tactics, which tend to have stronger correlations with outcomes (see Antonakis et al., 2011), the actor demonstrated appropriate nonverbal behavior to go with the relevant speech. The difference in the total instances of appropriate facial expressions and body language (evaluated on the basis of the studio videos) between the two conditions (i.e., 25 vs. 150) was highly significant ($\chi^2(1) = 89.29$, $p < 0.001$).[17]

The non-charismatic speech was a relatively solid speech with good use of rhetorical techniques and appropriate use of facial gestures and voice, though with a more subdued delivery in the use of body language. This approach was necessary to make the speech realistic and consistent with the underlying motivational content and so as to have a fair treatment comparison (Cooper & Richardson, 1986). Moreover, given the charitable context in which the work was conducted, we held the strength of moral conviction constant across the two speeches.

---

[16] We checked for inter-coder reliability. For the non-charismatic speech (n = 41 sentences), the coders agreed on 97.02% of the 369 coding events (i.e., 41 sentences * 9 categories). This level of agreement can be tested against chance agreement: $\kappa = .63$, $z = 12.28$, $p < 0.001$ (Landis & Koch, 1977). We found similar results for the charismatic speech (n = 48 sentences) with 95.60% agreement over 432 coding events: $\kappa = .74$, $z = 15.44$, $p < 0.001$. After having coded the speeches individually, the coders reconciled their codings on both speeches until reaching agreement. Appendix D reports the coded results for the two speeches.

[17] Specifically, the first coder coded 16 and 13 respective instances of use of appropriate facial expressions and body language in the non-charismatic speech; the second coder coded 10 and 11 instances (thus the means are 13 and 12 respectively). For the charismatic speech, the codings were 75 and 81 (first coder), and 71 and 74 (second coder), for means of 73 and 77 (rounded down).

For our initial subjective manipulation check we recruited 57 students from the University of Birmingham (situated in the same city in which we held the field experiment), whom we randomly assigned to watch a studio filmed video of the baseline ($n = 27$) or the charismatic ($n = 30$) speech. We used the idealized influence and inspirational motivation scales of the Multifactor Leadership Questionnaire (MLQ) to measure perceived charisma (Banks et al., 2017). In addition, we also included the General Leadership Impression (GLI) questionnaire, which taps into how prototypical the leader seemed to the participants (Cronshaw & Lord, 1987).[18] Following the advice of an anonymous reviewer, we performed a second subjective manipulation check using non-student participants who were more similar to the workers in our field experiment. For this purpose, we recruited 301 UK-based participants from the online platform Prolific Academic ($n = 152$ and $n = 140$, for the standard and the charismatic speech, respectively). In addition to the MLQ and GLI questionnaires, we also asked participants to directly rate the level of charisma of the speech, and included questions on perceptions of the work task and emotions associated with each speech.[19]

We perform OLS estimations (with robust standard errors) in which we regress indices for MLQ and GLI (created by taking the average of the questionnaires' items) on an indicator variable for the charismatic speech. The results for MLQ confirm that both samples of participants evaluate the speaker as more charismatic when delivering the charismatic speech, by 0.386 (2.667 vs. 3.053), p = 0.021) and 0.311 (3.828 vs. 4.139, p < 0.001) for student and online participants, respectively (where the scale ranged from 0 to 4). This result is further reinforced by the fact that—when asked directly—our online participants evaluated the level of charisma as significantly higher, by 0.842 (4.578 vs. 5.420, p < 0.001) on a scale from 1 to 7, in the charismatic speech than in the standard speech. The speaker was also seen as somewhat more prototypical for a leader (GLI) when giving the charismatic speech, but this effect is neither significant for the student sample (2.437 vs. 2.773, p = 0.164) nor for the online sample (3.722 vs. 3.870, p = 0.221). Finally, regarding task perceptions and emotions our online test reveals that the charismatic speech seems to neither make the work task appear more meaningful, nor to induce different emotions. Appendix Tables F1 and F2 report the full results for the student and online manipulation checks, respectively.

---

[18] Appendix F reports the instructions, survey instrument and implementation details for the subjective manipulation checks.

[19] The questions on task meaningfulness were taken from the RAND Corporation American Working Conditions survey and the Pew Research Center State of American Jobs survey.

3.6 Predictions: Effects of Incentives and Charisma on Worker Effort

We apply the framework from Section 2.2 to analyze our field experiment. In essence, in this situation, a leader aims at motivating workers to produce a simple, one-dimensional output.[20] For notational simplicity, we assume that the worker's action a directly corresponds to produced output and drop the worker specific subscript i. The material benefit m corresponds to the salary. Workers are either hired under a fixed wage contract in which they receive the wage w irrespective of their performance ($m = w$), or they are hired under a performance pay contract in which they get—in addition to the fixed wage w—a piece rate p for each output unit that they produce beyond a certain threshold output t ($m = w + p \cdot \max[0, a - t]$). The worker's intrinsic utility $n > 0$ depends on the worker's output a ($\partial n/\partial a > 0$, $\partial^2 n/\partial a^2 \leq 0$, $\lim(a \rightarrow 0)\ \partial k/\partial a = \infty$, and $\lim (a \rightarrow \infty)\ \partial i/\partial a_i = 0$)) and the leader's charisma c ($\partial n/\partial c > 0$ and $\partial^2 n/\partial a \partial c > 0$).[21] We use a standard, convex effort cost function k ($\partial k/\partial a_i > 0$, $\partial^2 k/\partial a_i^2 > 0$, $\lim (a_i \rightarrow 0)\ \partial k/\partial a_i = 0$, and $\lim (a_i \rightarrow \infty)\ \partial i/\partial a_i = \infty$).

The slope of an indifference curve in the money-output space is given as follows:

$$\partial m/\partial a = \partial k/\partial a - \partial n/\partial a.$$

Our assumptions on n and k imply that this slope is negative for low levels of output (because initially the increase in intrinsic satisfaction from additional output dominates the cost increase) and positive for high levels of output (because the cost increase becomes dominant).

[Figure 2 about here]

Figure 2 shows examples of indifference curves in the worker's benefit-output space for both low leader charisma (c_L, green lines) and high leader charisma (c_H, red lines). In the baseline and charisma conditions, the salary corresponds to the fixed wage w, so that the set of points that the worker can reach consists of a flat line (blue line). In the piece-rate condition the set of reachable points includes a kink (kinked blue line), because the salary starts to increase once output surpasses the threshold t ($m = w + p \cdot \max[0, a - t]$). The worker's aim is to reach the highest indifference curve possible. In the baseline condition the utility maximizing output

---

[20] One might worry about a quantity-quality trade-off. However, workers are aware that quality may be inspected when they return their completed envelopes for payment. In the results section we report quality measures that show that workers in the charisma and incentive treatments did not increase quantity at the expense of quality.

[21] We assume that $n > 0$, because the speeches focus on the positive consequences of working hard rather than the negative effects of shirking. However, our hypotheses do not depend on this assumption (see also Section 2.2). We abstract from the fact that a worker's intrinsic benefit may also depend on other workers' performance. Allowing for strategic complements would considerably complicate the exposition without generating decisive new insights. We discuss strategic complements when we introduce the public goods game in the next section.

choice is a_B. The corresponding levels in the charisma condition and the piece rate condition are a_C and a_P. Thus, both incentives and charisma increase performance.

Performance pay allows the worker to reach money-output combinations that are not available under the fixed wage. If the marginal cost of effort ($\partial k/\partial a$) is not yet too high when output reaches threshold ($a = t$), the worker can increase utility by increasing output to a level above the threshold (the additional income more than compensates the marginal cost). Obviously, the piece rate may have no effect for some workers—if the marginal cost of effort is already high at the threshold, the indifference curves are very steep and such workers will be unable to increase their utility by working harder. However, as long as there are some workers in the population for whom the piece rate is attractive, overall output will strictly increase.

FE-Hypothesis 1: *The output in the piece rate condition is higher than the output in the baseline condition.*

The charisma effect is a direct consequence of our assumption that leader charisma increases the marginal intrinsic satisfaction from working harder ($\partial^2 n/\partial a \partial c > 0$). The positive cross derivation flattens the indifference curves (the salary increase necessary to keep utility constant when effort increases gets smaller) so that the worker's optimal output level increases.

FE-Hypothesis 2: *The output in the charisma condition is higher than the output in the baseline condition.*

Without further assumptions, our model does not make predictions regarding the relative effect sizes for charisma and incentives, because this comparison depends on the incentive intensity, the size of the charisma impact, and the functional forms of *n* and *k*. To make the comparison of effect sizes in the two treatments informative, we implemented the highest piece rate that still makes business sense in our environment. In this sense, we test our charisma manipulation against the highest piece rate that can be used without losing money.

3.7   Results

We first provide descriptive statistics before analyzing the results with regression analysis.

Descriptive statistics

Table 1 presents descriptive statistics. On average, workers in the baseline condition return 230.9 completed envelopes so that the average cost per letter amounts to GBP 0.123 (GBP 28.4 / 230.9). In line with *FE-Hypotheses 1 and 2*, both the piece rate treatment and the charisma treatment substantially increase the number of envelopes. In the piece rate condition workers

return, on average, 277.7 letters (an increase in average output of 20.2% relative to the baseline condition). However, because workers who complete more than 220 envelopes receive a piece rate of GBP 0.12 for each additional envelope, the workers' average earnings increase from GBP 28.4 to GBP 37.2. As a consequence, there is a small increase in the average cost per letter to GBP 0.134 (an increase of 8.8% relative to the baseline). Non-parametric rank-sum tests indicate that the increase in performance is statistically significant ($p = 0.025$), whereas the increase in the cost per letter is not ($p = 0.580$).

[Table 1 about here]

In the charisma condition, workers return, on average, 271.2 letters. This performance is very similar to that in the piece rate treatment—a non-parametric rank-sum test reveals no statistically significant difference ($p = 0.802$). Relative to the baseline condition, performance in the charisma treatment is 17.4% higher. Because earnings were unaffected by the charisma manipulation, the higher number of completed letters also results in a substantial reduction of the cost per letter to GBP 0.105 (a cost decrease of 14.8% relative to the baseline condition). Non-parametric rank-sum tests indicate that both effects are statistically significant ($p = 0.017$ in both cases, as the tests are substantively identical).[22] This measurement of the cost reduction does not consider that the creation of a charismatic speech itself may also come at an additional cost. However, our data allow us to determine how much money could have been spent on the charismatic speech without increasing the overall cost of the donation campaign. For a planned number of 30,000 letters to overall cost reduction amounts to about GBP 540. Thus, as long as the cost of increasing the charisma of a five-minutes motivation speech does not exceed this amount, using charisma remains cost effective in our setting.[23]

Figure 3 summarizes the main effects of our treatment conditions graphically and Figure 4 reports the distributions of the number of letters completed across treatments.

[Figures 3 and 4 about here]

As mentioned in Section 3.4, we not only recorded the number of completed envelopes, but we also performed quality checks to investigate, for example, whether workers in the monetary bonus treatment were tempted to compromise on quality. Overall, our quality inspection yields a series of 20 check items, each of which was judged as either a PASS or a

---

[22] The variance of the distribution of the number of letters stuffed is not different between the two treatments (F-test for the homogeneity of variances, $p = 0.115$).

[23] It is important to keep in mind that there was a speech in all our treatments. Thus, the relevant cost is therefore not the cost of producing the speech per se, but rather the cost of making the speech more charismatic.

FAIL.[24] Out of these 20 items we generate a simple index that calculates the relative frequency of passed items. As a second measure we use an indicator variable for perfect quality which is unity if all items checked have been passed and zero otherwise.[25]

The evidence presented in Table 1 suggests that workers performed their work quite carefully in all three conditions. In the baseline conditions workers passed on average 86% of the quality items. In the piece rate and charisma conditions the corresponding numbers are 88% and 90%, respectively. Non-parametric rank-sum tests indicate that there are no statistically significant differences across treatments (baseline vs. piece rate: $p = 0.548$, baseline vs. charisma: $p = 0.287$, piece rates vs. charisma: $p = 0.670$). Samples of perfect quality were somewhat more frequent in the piece rate condition (30%) and the charisma condition (32%) than in the baseline condition (21%). However, none of these differences is statistically significant, again using non-parametric rank-sum tests (baseline vs. piece rate: $p = 0.389$, baseline vs. charisma: $p = 0.253$, piece rate vs. charisma: $p = 0.825$). We admit that these results need to be interpreted with caution, because our quality data is not very precise (we only had a chance to inspect three envelopes per worker), but—according to the data that we have—the increase in completed letters observed in the piece rate and charisma conditions did not seem to come at the cost of decreased quality.

Table 1 also displays the worker's self-reported information regarding time spent on the task and whether they had help from friends or family. Workers in the piece rate and charisma conditions report to have worked about 20 minutes longer than workers in the baseline and seem also somewhat more likely to have had help from friends or family. However, neither of these differences is statistically significant (non-parametric rank-sum tests for time spent on task: baseline vs. piece rate: $p = 0.185$, baseline vs. charisma: $p = 0.613$, piece rate vs. charisma: $p = 0.449$; Fisher's exact test for help from friends or family: baseline vs. piece rate: $p = 0.488$, baseline vs. charisma: $p = 0.533$, piece rate vs. charisma: $p = 1.000$).

---

[24] In particular, for each worker we randomly drew three envelopes and checked whether the material in the envelope was complete and in the right order, whether the handwritten code on the donation form was correct, and whether the letter was neatly folded and free of printing faults. In addition, we also checked whether the returned envelope bundles were nicely assembled (no corners bent), consistently labeled, and correctly secured with rubber bands. Finally, we recorded whether the envelope bundles were returned in the designated boxes and whether they were neatly arranged.

[25] As a consequence of a coordination problem, we failed to collect complete quality information for eight of our 106 participants. In six cases, we performed quality checks on a subset of items, while for the remaining two workers we have no quality information. For the former set of workers, we compute the quality indicators on the subset of checked items. In the remaining two cases (one in the baseline and one in charisma), we code the quality indicators as missing.

Regression Analysis

Table 2 reports the results of a series of regression estimations in which we further explore the statistical significance of the treatment effects reported above. In all estimations, we regress the outcome variable on indicator variables for the piece rate condition and the charisma condition. Our workers performed the work task alone at their homes, and interactions between workers during the instruction sessions were minimal (the sessions were tightly timed and workers had essentially no opportunity to talk). Our main estimation therefore treats each worker as an independent observation and we report robust standard errors (Panel A). However, because we cannot exclude that subtle forms of interaction may have taken place during the instruction sessions (e.g. non-verbal exchanges, etc.), we also include additional specifications, where we use a conservative approach and cluster standard errors at the session level, to allow for the possibility of intra-group correlation (Panel B). We use the Moulton method to correct for the small number of clusters (Angrist and Pischke, 2008).

In Column (1) we consider the sample of workers who completed the work task and evaluate treatment effects on the number of completed letters per worker through an OLS estimation. The constant corresponds to the average number of completed envelopes in the baseline condition (231 letters), whereas the indicator variables capture the corresponding increase in each of our two treatment conditions, separately. The estimation results reported in Panel A confirm that piece-rate and charisma treatments increase output by 47 and 40 letters, respectively, and that both these treatment effects are statistically significant ($p = 0.026$ and $p = 0.040$, respectively). An F-test confirms that there is no significant difference between the two treatment effects ($p = 0.738$). Column (2) performs the same estimation, but uses a data set that includes the workers who dropped out of the experiment after the treatment as observations with zero output. Because the dropouts are concentrated in the baseline and piece rate conditions, we observe a decrease in the constant (i.e., a lower average output of 213 instead of 231 letters in the baseline) and a larger and highly significant charisma coefficient ($p = 0.007$) which now indicates an increase of 58 letters. The piece rate coefficient remains similar (48 letters), but is slightly less significant ($p = 0.058$). Again, the treatment effects of piece rate and charisma do not significantly differ ($p = 0.625$).

In Column (3) we perform an OLS estimation to examine treatment effects on the cost per letter (measured at the worker level). The constant reveals a cost per letter of GBP 0.146 in the Baseline. The regression reveals that the cost reduction observed in the charisma condition (GBP 0.028) is statistically significant ($p = 0.048$), whereas the small cost increase observed in the piece rate condition (GBP 0.006) is not ($p = 0.760$).

[Table 2 about here]

Columns (4) and (5) show estimations in which we regress our quality measures on indicator variables for the piece rate condition and the charisma condition. Column (4) uses the quality index as the dependent variable, whereas Column (5) uses the indicator variable for perfect quality. Both estimations show that neither the piece rate treatment nor the charisma treatment has a significant effect on the quality of the completed envelopes.[26]

Panel B replicates the analysis using a more conservative estimation approach that accounts for within-session interdependence. Not surprisingly, standard errors are slightly higher when clustered at the session level, but the significance levels do not change much. The significance level of our main variable of interest (charisma) drops in Column (1) but the p-value remains very close to the 5% threshold (p = 0.056). Thus, our results hold up and are only slightly weaker if we apply the conservative approach of clustering at the session level.

In Table 3 we add control variables to the regressions on our main outcomes and consider additional outcomes collected through the worker questionnaire. Because these latter outcomes are self-reported and some subjects failed to respond to these questions, these results need to be interpreted with caution. As in the previous table, we again report estimations with robust standard errors (Panel A) and estimations with standard errors clustered at the session level (Panel B). Column (1) adds workers' age and gender as control variables to the estimation in which we regress the number of completed letters per worker on indicator variables for the piece rate condition and the charisma condition (note we lose one observation, and hence statistical power, because one subject failed to report demographic information). The treatment effect of the piece rate condition reported in Panel A remains almost unchanged and significant (p = 0.035). The treatment effect of the charisma condition is slightly lower, at 37 letters, but remains marginally significant (p = 0.071). Column (2) repeats the same exercise for the cost per letter. As before we do not find a significant effect for the piece rate (p= 0.728), but charisma marginally significantly reduces the cost per letter (p = 0.089). Thus, using age and gender to

_____

[26] We compute Minimum Detectable Effects, or MDEs (using the conventional 5% level of statistical significance and 80% power level), to make comparisons between our study and the ones available in the literature (McKenzie and Ozier, 2019). We follow Haushofer and Shapiro (2016) and report MDE only for non-significant parameters to distinguish between lack of power from lack of effect. The ex-post MDE is of 0.031, 0.095 and 0.308 for the piece-rate treatment relative to the baseline in terms of cost per letter, the quality index and perfect quality, respectively; and of 0.084 and 0.288 for the charisma treatment relative to the baseline in terms of the quality index and perfect quality. These figures suggest that any insignificant effect that we find on cost per letter and the quality index are not due to lack of power, but are instead the result of the small size of treatment effects.

control for possible imperfections in randomization across treatments leaves the pattern of our main results largely unchanged.[27]

[Table 3 about here]

Columns (3) to (6) confirm that neither the time spent on the task nor the frequency with which workers get help from friends or family are significantly affected by our treatment manipulations. These results hold independently of whether or not the estimation controls for age and gender. Panel B shows that our main results in Table 3 are robust to clustering standard errors at the session level.

4   Laboratory Experiments

Our field experiment provides evidence that enhancing the degree of charisma in a leader's motivational speech can have large effects on individual follower effort—as large as those of a substantial increase in incentive pay. However, a better understanding of the effects of charisma requires more data to, for instance, test varying conditions under which it is and is not effective. Moreover, the results of our field experiment raise a number of questions regarding what factors are important for charisma to have a positive influence. Such questions are hard to study in the field, because a thorough investigation of the limits and mechanisms underlying the effects of charismatic leadership require a large number of treatments and the elicitation of measures that are hard to obtain naturally outside of the laboratory (e.g., data on beliefs). Thus, the ability to study and unpack the mechanisms underlying charisma's effects require a more practical paradigm for studying its influence. To advance these objectives, we complement our field experiment with a series of laboratory experiments that test the motivating effect of charisma in a different context, using a different outcome—voluntary contributions to a public good. Moreover, this approach lends itself to easier replication and further experimentation.

Our design builds on a linear public goods game with fixed groups of four players who interact for 10 periods. We chose the public goods game, because it is the most widely employed paradigm for experimentally investigating collective action. Moreover, due to its inherent tradeoffs between myopically selfish behavior and pro-social and efficient cooperation, the

---

[27] Using the covariates reduces the sample, which decreases statistical power and may bias parameter estimates. We therefore also performed estimations in which we use a maximum likelihood estimator for handling missing data. Under this approach, the effect of charisma is significant at the 5% level (see Appendix E Table 1 for a more detailed explanation and results across a range of dependent variables). In addition, in Appendix E we repeat the analysis of Table 2 using a Tobit model with censoring at 350 letters. These regressions account for the fact that workers could not complete more than 350 letters. Because only two workers (both in the charisma condition) completed the full 350 envelopes, the estimation results do not change much.

game is useful to investigate whether a leader's communication can produce the kinds of transformational influences often discussed in the organizational literature (House, 1977).[28] However, it is important to note that the laboratory public good context we study differs in many ways from the setting and behavior in our field experiment. Rather than seeking to produce as close to a directly replication of the field context in the laboratory as possible, our aim is to investigate the effects of charisma in an alternative—and widely studied—laboratory setting that contains some, but not all, features of the field setting.

We conducted three laboratory experiments in which participants are given a speech before they play a public goods game. To begin with a challenging test for charisma effects, we started with a simple setting in which we simply added a motivational speech to a typical public goods experiment. Our remaining two experiments introduce additional features present in the field experiment that might enhance the effects of charisma. Thus, we attempt to partially bridge the gap from highly abstract settings where charisma might be unlikely to have effects to our rich field setting where it seems to have a large effect.[29]

In each experiment, participants play the same public goods game for 10 periods. Before the first contribution round subjects in the treatment conditions get to see a video, in which an outsider (a professional actor), who has no stakes in the game, delivers a speech intended to motivate the players to contribute to the public good. In the speech, the actor explains the social dilemma and emphasizes that the group benefits if all players act in a "morally correct" manner and contribute to the shared project. In the instructions participants are told that they will watch a video, but they do not receive any information about the purpose and origin of the message before actually watching the video. In addition to the incentivized contribution decisions, we also elicit players' beliefs about the average contribution of the other members in their group.

We are aware that it is unusual to provide participants with direct guidance on how to behave in an experiment. In many cases, such a procedure would (correctly) be deemed inappropriate because it is predestined to trigger undesired experimenter demand effects (Zizzo, 2010; De Quidt et al., 2018). However, our purpose is to investigate the effectiveness of

---

[28] For other studies that use public goods games to study the effects of leadership, see Van Vugt & De Cremer, 1999; Güth et al., 2007; Dal Bo & Dal Bo, 2014; and Kosfeld and Rustagi, 2015.

[29] We also conducted a separate study on charismatic communication, in which subjects playing similar public good games received written messages from other subjects encouraging them to contribute. The messages were written by students from the same subject pool, who received a share of the total amount contributed. In a standard treatment, messages were simply the ones written by these subjects, while in an "enhanced charisma" treatment the same messages were re-written by the research team, by adding CLTs. This manipulation of charisma had no effect on contributions. This study differs in many respects from the laboratory and field experiments we report here and represents a substantively different investigation into the effects of charismatic communication. Nevertheless, as it sheds light on the limits of charismatic communication, we briefly summarize the design and results in the appendix of this paper (see Appendix G).

different kinds of statements that make substantively similar requests, so a demand effect is always present in most comparisons we make with our data; that is, one interpretation of our studies is that we estimate the treatment effect of charismatic versus non-charismatic demand effects. In the world of practice, requests from leaders to exert a costly collective action may be thought of as a form of demand effect; in this sense, our experiments test whether charismatic leadership tactics can be used to create a significantly stronger demand effect.

4.1 Laboratory Experiment 1: A Stark Test

Design and Treatments

Participants are assigned to groups of four players that remain fixed for all 10 periods. In every period, each player receives an endowment of 20 points. Players can freely allocate their endowments between their personal account and a common pool. Each point in the personal account directly adds to the player's payoff. Points that are contributed to the common pool yield a marginal per capita return (MPCR) of 0.4 points (i.e., a total return of 1.6 points to the group). We can therefore write Player i's payoff π_i, as a function of her own contribution (c_i) and others' contributions (c_j), as:

$$\pi_i = 20 - c_i + 0.4 \sum_{j=1}^{4} c_j.$$

In Experiment 1 we implement three treatment conditions. We run two conditions with different versions of the motivational speech: a charismatic and a non-charismatic one, and compare them to a control condition without a speech. Both motivational speeches last about 4 minutes, have the same number of words, and are substantively very similar. Participants were made aware of the fact that all four members of the group in which they interact see the same speech. The speeches explain the intuition behind the tragedy of the commons, re-explain the payoff structure of the game and emphasize the social obligation of each player to act in the interest of the group (the full text and links to the videos are found in Appendix B). As in our field experiment, the difference between the two speeches is that the charismatic version makes extensive use of CLTs. The videos were displayed on each participant's individual computer screen.

Procedures

Subjects received CHF 10 for their participation, as well as their accumulated earnings across the 10 periods of the experiment. Points were converted into money at the exchange rate of 20 points = CHF 1 (approximately, USD 1).

23

We conducted all sessions at the University of Lausanne. Materials and instructions were in French.[30] We ran 24 sessions in total: 9 sessions for each of the two speech treatments and 4 sessions for the control treatment. We recruited 24 participants for each session, but due to no-shows we only had 20 or 16 participants in some sessions. In total we had 436 participants in the experiment. All participants were students from the University of Lausanne, EPFL Lausanne, or EHL Lausanne. We used the software ORSEE (Greiner 2015) for recruitment. All sessions were run between November 2015 and March 2016. The experiment was programmed and conducted with z-Tree (Fischbacher, 2007).

Manipulation checks

As in our field experiment, we conducted objective and subjective manipulation checks to ensure that subjects from the same pool of students perceived the difference in charisma. Two expert coders coded one of the speeches independently; then they coded the other speech. Initial agreement on the charismatic speech, which had 40 sentences (thus 40 * 9 categories = 360 coding events) was 96.83% (κ = .71, z = 14.31, p < 0.001). The non-charismatic speech, which had 32 sentences (thus 32 * 9 categories = 288 coding events) had an agreement of 99.70% (the κ was undefined because one coder found only one charismatic tactic in the whole speech, whereas the other coder found none). After reconciling their ratings, the number of charismatic leadership tactics used as a proportion of total sentences in the non-charismatic speech was 0%; that of the charismatic speech was 50% (i.e., 20 tactics over 41 sentences). This difference in proportions is highly significant (z = 4.64, p < 0.001, see Koopman, 1984).[31] The actor demonstrated appropriate body language for the level of verbal tactics used.

For the subjective manipulation check, we recruited 60 students from the University of Lausanne, EPFL Lausanne, or EHL Lausanne, whom we randomly assigned to watch either the standard (non-charismatic, *n* = 31) or the charismatic (*n* = 29) speech. We then administered the same questionnaires that we used in the field experiment manipulation check, and analyzed the results using the same methodology. The OLS estimation results indicate that the actor is perceived as more charismatic and more prototypical for a leader when he gives the charismatic speech (p < 0.001 both for the MLQ index (2.091 vs. 3.080) and for the GLI index (1.555 vs. 2.842), see Appendix Table F1).

---

[30] For a translated sample of our instructions for participants in the laboratory experiments see Appendix H.
[31] Appendix D reports the coded results for the two speeches.

Predictions: Effects of Charisma on Contributions to a Public Good

We apply the theoretical framework from Section 2.2 to a public good game with N players. Each player has an endowment of E units that can either be used for private consumption or can be contributed to the public good. We denote player i's action, $a_i \in [0, E]$, as the player's contribution to the public good. Each unit that a player contributes creates a marginal per capita return, α, for each of the N players. Player i's material benefit therefore corresponds to $m = E - a_i + \alpha \Sigma_i a_i$. To obtain a closed-form solution, we assume that the intrinsic benefit is given by the expression: $n = c\,(\gamma + \delta \Sigma_{j \neq i} a_j) \ln a_i$, where $\gamma > 0$ and $\delta > 0$.[32] This function has the following properties (assuming that the leader's perceived charisma, c, is always positive): i) the intrinsic benefit increases with a decreasing marginal rate in the follower's action a_i, ii) the marginal intrinsic benefit of action a_i is positive even if all other players do not contribute to the public good, and iii) the marginal intrinsic benefit of action a_i increases in the contributions of other players (and increasingly so when leader charisma is high), reflecting a tendency for individuals to exhibit conditional cooperation in linear public good games (see, e.g., Fischbacher et al. 2001). Player i's utility is thus given by:

$$u_i = E + \alpha \sum_i a_i + c\left(\gamma + \delta \sum_{j \neq i} a_j\right) \ln a_i - a_i.$$

Players face a social dilemma; that is, it is socially efficient to contribute the full endowment to the public good, but it is individually rational to keep the full endowment for private consumption ($\alpha < 1 < \alpha N$).

Considering that individual contributions to the public good are bounded above by the endowment E, Player i's best response function is given by:

$$a_i = min\left[E, \frac{c}{1 - \alpha}\left(\gamma + \delta \sum_{j \neq i} a_j\right)\right].$$

Assuming that all players have the same utility function, symmetry ($a_i = a_j \; \forall \; i, j$) implies the following equilibrium contribution to the public good for every player:

$$a_i^* = \begin{cases} \dfrac{c\gamma}{1 - \alpha - c\delta(N-1)}, & \text{if } 0 \leq \dfrac{c\gamma}{1 - \alpha - c\delta(N-1)} < E, \\ E, & \text{otherwise.} \end{cases}$$

---

[32] This specific functional form implies that $n > 0$. However, adding a negative constant to the function would not affect the analysis so that our hypotheses hold irrespective of the sign of intrinsic utility (see also Section 2.2).

In this setting, an increase in the leader's charisma has two effects. On one hand, an increase in charisma directly increases the follower's intrinsic motivation. On the other hand, there is an indirect effect stemming from the strategic complementarity in follower contributions. The overall impact of an increase in charisma on contributions to the public good is unambiguously positive as long as equilibrium contributions are inferior to the endowment:[33]

$$\frac{\partial a_i^*}{\partial c} = \frac{\gamma[1 - \alpha - c\delta(N-1)] + c\gamma\delta(N-1)}{[1 - \alpha - c\delta(N-1)]^2} > 0.$$

[Figure 5 about here]

Figure 5 graphically shows the impact of an increase in charisma on equilibrium contributions for a 2-player version of the model.

LE-Hypothesis 1: *In Laboratory Experiment 1, contributions to the public good are higher in the charisma condition than in the no charisma condition. Moreover, contributions in both conditions with speeches are higher than contributions in the baseline condition.*

Results

Panel A of Figure 6 shows average contributions and average beliefs over time for the three treatments. The figure reveals that directionally, *LE-Hypothesis 1* holds. Average contributions in both treatments with motivational speeches are higher than average contribution in the control treatment in every period, and average contributions in the charisma treatment are consistently higher than those in the no charisma treatment. Averaged over all periods, participants contribute 7.9 points in the control treatment, 10.1 points in the no charisma treatment and 11.2 points in the charisma treatment. Contributions in the charisma treatment are therefore roughly 11 percent higher than those in the no charisma condition and about 42 percent higher than those in the control condition. The pattern of beliefs is very similar to that of contributions. Participants tend to slightly overestimate others' contributions (overall average beliefs exceed average contributions by roughly 1 point in all three treatments), but beliefs are highest in the charisma condition and lowest in the control.

[Figure 6 about here]

---

[33] We think of our control condition without a speech as a situation in which leader charisma c is very small (but positive), so that players still also have some intrinsic motivation to contribute to the public good. This view seems justified given that the written instructions to the public goods game also include (very factually delivered) portions highlighting the tension between individual profit and social welfare, which are more persuasively presented in the leader's speeches.

In Table 4 we report a regression analysis of the treatment effects on contributions (Panel A) and beliefs (Panel B). We perform simple treatment comparisons in which we regress the dependent variable on treatment dummies (see columns (1) and (3)) and estimations that identify separate time trends for the treatments (see columns (2) and (4)). Because our primary interest is in differences between the no charisma and charisma conditions, we use the former as the omitted category. We estimate both specifications first with OLS and then with Tobit estimators (to account for the fact that contributions are censored at the maximum level of 20 points) with a cluster-robust estimate of the variance.

Our estimations reveal that the charisma treatment increases contributions by 1.1 points relative to the average contributions of 10.1 points in the no charisma condition (see the constant), while contributions in the control treatment are lower than in the no charisma condition by 2.2 points. However, it turns out that only the difference between the charisma treatment and the control treatment is significant (F-Test: Charisma = Control, $p = 0.023$ (OLS), $p = 0.015$ (Tobit)). The difference between the no charisma treatment and the control condition is not significant ($p = 0.144$ (OLS), $p = 0.110$ (Tobit)). Most importantly for our purposes, the difference between the two treatments with different speeches is also not significant ($p = 0.353$ (OLS), $p = 0.385$ (Tobit)). Thus, we cannot reject the null hypothesis that the charismatic and the non-charismatic speeches have the same effect on the participants' contribution behavior. There is substantial cross-group variance in the data, which leads to large standard errors. The fact that behavior differs strongly across groups is not surprising, because conditional cooperation implies that the group composition plays a huge role for the development of average contributions over time (e.g., Fischbacher and Gächter, 2010).

[Table 4 about here]

4.2 Laboratory Experiment 2: Including Charitable Contributions

Design and Treatments

Our first laboratory experiment found a modestly positive—but statistically insignificant—treatment effect of increasing the charisma of a leader's message on contributions. This finding suggests that the benefits of charisma in our field experiment might be less pronounced in other contexts, where many factors present in the field setting are absent. It might be that the absence of one or more of these factors contributed to the null result, or it may alternatively be that the benefits of charisma are so ephemeral as to be hard to reproduce. To shed light on these competing possibilities, our next two experiments add features present in our field experiment to the abstract public goods game used in the first laboratory experiment.

First, we consider the fact that our laboratory design has a diminished moral dimension relative to the field experiment. The public good game involves a moral element, but contributing to a common project that benefits fellow students may be less morally charged than helping hospitalized children. Our second laboratory experiment tests whether the charisma manipulation is more powerful when including this additional element of morality.

The game and the payoff structure remain the same as in our first laboratory experiment. The only additional feature is that in experiment 2 each point that a player contributes to the public good also generates a donation of CHF 0.05 to the children's hospital in Lausanne. Because the experiment is played for 10 periods and each player can contribute up to 20 points per period, an individual participant can donate up to CHF 10 throughout the experiment.

Because our primary interest is in the comparison of the effects of the two speeches and the previous experiment revealed that a charismatic speech increases contributions significantly relative to the control condition, we only conducted two treatments: one with a charismatic speech and one with a non-charismatic speech. The speeches we used combine elements of the speeches used in the first laboratory experiment with elements of the speeches used in the field experiment (the full text of the two speeches and links to the videos are found in Appendix B). As in the previous laboratory experiment participants were made aware that all members of a group watch the same video.

Procedures

We conducted 13 sessions in total—7 for the charisma treatment and 6 for the no charisma treatment. We recruited 24 participants for each session, but because of no-shows some sessions only had 20 participants (we had 300 participants in total). The subject pool, the recruitment process and the software used to program and run the experiment were the same as in the first laboratory experiment.[34] All sessions were run from November to December 2016.

Manipulation check

We again conducted objective and subjective manipulation checks of the two speeches. We followed the same procedures, although using different raters from the same subject pool in the subjective manipulation check. Given that the speech was very similar to the previous speech, the objective coders coded the speeches together until reaching full agreement. The coders indicated that for the non-charismatic speech the proportion of sentences including charismatic tactics was 14.29% (i.e., 4 tactics over 29 sentences); that of the charismatic speech was 81.82%

---

[34] For a translated sample of our instructions for participants in the laboratory experiments see Appendix H.

(i.e., 27 tactics over 34 sentences). This difference in proportions was significant ($z = 5.19$, $p < 0.001$, Koopman, 1984).[35] We also performed the same subjective manipulation check as in Laboratory Experiment 1 (using 58 students, whom we randomly assigned to watch either the non-charismatic ($n = 29$) or the charismatic ($n = 29$) speech). Our OLS estimations produce very similar results as for the previous speeches. Once again, the actor is perceived as more charismatic and more prototypical for a leader when he performs the charismatic speech ($p = 0.024$ both for the MLQ and for the GLI index, see Appendix Table F1).

Predictions: Effects of Adding a Charitable Donation to the Public Goods Experiment

The intent behind including a charitable donation was to increase the importance of the moral dimension. The addition of the donations does not influence followers' pecuniary returns from their own or others' contributions. However, stronger moral implications may imply that followers experience a stronger intrinsic motivation to contribute to the public good. In our model, the importance of the intrinsic benefit is determined by the parameter γ. An increase in γ has two implications: i) a higher γ increases equilibrium contributions for a given level of charisma: $\partial a_i^*/\partial\gamma > 0$, and ii) the increase in equilibrium contributions is larger the higher charisma: $\partial^2 a_i^*/\partial\gamma\partial c > 0$. These effects yield the following hypothesis:

LE-Hypothesis 2: *In Laboratory Experiment 2, contributions to the public good are higher than in Laboratory Experiment 1 in both the charisma condition and the no charisma condition. However, the increase in the charisma condition is larger so that the difference between the charisma condition and the no charisma condition is larger in Laboratory Experiment 2 than in Laboratory Experiment 1.*

Results

Panel B of Figure 6 shows the consequences of adding the charitable donation to the public good game. In line with the first part of *LE-Hypothesis 2*, the link of the contribution to a donation increases average contributions in both treatments relative to the first experiment, from 10.1 to 13.6 points in the no charisma treatment and from 11.2 to 13.9 in the charisma treatment. However, these numbers also imply that the second part of *LE-Hypothesis 2* is not supported by the data, because the difference in contributions in the charisma condition (13.9) and the no charisma condition (13.6) is very small.

---

[35] Appendix D reports the coded results for the two speeches.

Table 4 shows, in models 5 through 8, the same within-experiment regressions for contributions and beliefs that we discussed for the first laboratory experiment. The small positive charisma effect (0.320 (OLS), 1.309 (Tobit)) is not statistically significant in either estimation ($p = 0.774$ (OLS), $p = 0.583$ (Tobit)). The regressions also reveal that—as in the first experiment—the effects on beliefs are very similar to the effects on contributions.

The results of the second experiment indicate that a morally laden environment alone does not guarantee that a more charismatic speech has a large effect on behavior, despite the fact that the strengthened moral dimension appears to increase the motivation to contribute.

4.3   Laboratory Experiment 3: Collective Viewing

Design and Treatments

A potentially important difference between our field experiment and the first two laboratory experiments is that, in the field experiment, the speech was given to participants in groups, whereas participants in our laboratory experiments watched the video individually in an isolated cubicle. This difference might be relevant if collectively perceiving the significance of the moral mission strengthens a collective sense of identity (Hogg, 2001; Turner et al., 1994) and/or if people respond positively to charisma because they think others will do so (Hermalin, 2017).

In our third laboratory experiment, participants within a session watch the video as a group in front of a large TV. Specifically, after reading the instructions and answering the control questions people move to a waiting room located right in front of the lab where they watch the video of either a charismatic or non-charismatic speech together. Importantly, *there was no verbal interaction between participants at any point in time*. The participants were instructed to be silent on the way to the TV, during the video, and on the way back to the laboratory, and were accompanied by an experimenter during this time.[36] Except for the joint viewing experience the experiment was identical to Laboratory Experiment 2.

Procedures

We ran 12 sessions with 228 participants in total. We conducted 6 sessions for each of the two treatments. We recruited 20 participants for each session, but due to no-shows some sessions only had 16 participants and one session only had 12 participants. All sessions took place in

---

[36] Note also that any opportunities for non-verbal communication between participants during this time were the same across conditions.

April and May 2017. Again, the subject pool, the recruitment process and the software used were the same as in the other two experiments.[37]

Predictions: Effects of Group Watching

The idea behind letting the participants watch the speech as a group was to intensify the perception of a collective endeavor. The stronger group perception might increase the complementarity between the contributions of the different followers in the group. In our model the complementarity between follower actions is captured by the parameter δ. The effects of an increase in δ are similar to those of an increase in γ (see the previous section). Our model suggests that the group watching procedure should increase contributions for a given level of charisma ($\partial a_i^*/\partial\delta > 0$) and the increase should be particularly strong when charisma is high ($\partial^2 a_i^*/\partial\delta\partial c > 0$). According to the model, these effects should be enhanced by the fact that the complementarity between actions reinforces the effect of charisma.

LE-Hypothesis 3: *In Laboratory Experiment 3, contributions to the public good are higher than in Laboratory Experiments 1 and 2 in both the charisma condition and the no charisma condition. However, the increase in the charisma condition is larger so that the difference between the charisma condition and the no charisma condition is larger in Laboratory Experiment 3 than in Laboratory Experiments 1 and 2.*

Results

The data collected in the third experiment are largely in line with *LE-Hypothesis 3 and* suggest that the experience of watching the video as a group makes a large difference for the impact of the charisma manipulation on behavior in the public good game. Panel C of Figure 6 shows that in the third experiment participants in the charisma condition consistently contribute substantially more to the public good than participants in the no charisma condition. Across all periods of the experiment, average contributions amount to 16.1 points in the charisma condition and to 13.6 points in the no charisma condition. This difference corresponds to an increase of about 19 percent, very similar to that observed in our field experiment. Panel A of Table 4, columns 9 and 11 confirm that in the third laboratory experiment the positive effect of charisma on contributions is statistically significant and provides supports for *LE-Hypothesis 3* (p = 0.034 (OLS), p = 0.023 (Tobit)). Columns 10 and 12 add time trends and interact them with the charisma treatment. The (significantly) negative time trend in combination with the

---

[37] The results from the Experiment 2 manipulation checks also apply to Experiment 3, because the speeches were exactly the same in the two experiments.

(non-significantly) positive interaction effect suggests that contributions tend to decline more slowly in the charisma treatment. If we only use first-period data, the effect of charisma on average contributions is rather small (16.7 points (charisma) vs. 15.6 points (no charisma), $p = 0.124$ (OLS), $p = 0.083$ (Tobit)), but the effect on contributions tends to become larger over time. Period-specific tests of linear combinations of coefficients in the OLS model in column 10 (Charisma + (Period x Charisma) x Period) reveal that the impact of charisma on contributions only becomes significant from Period 3 onwards ($p = 0.043$ in period 3).

Average beliefs increase from 13.7 points in the no charisma condition to 15.8 points in the charisma condition; columns 9 and 11 of Panel B of Table 4 reveal that this effect is marginally significant ($p = 0.051$ (OLS), $p = 0.078$ (Tobit)). The dynamics of beliefs are similar to those of contributions (see columns 10 and 12). If we only look at initial beliefs in period 1 (before any contribution was made), we do not find a significant difference in beliefs across treatments ($p = 0.175$ (OLS), $p = 0.109$ (Tobit)). We interpret this observation as an indication that the collective viewing experience is more likely to operate through the creation of a collective identity rather than through the belief channel. In conjunction with the absence of a large effect of charisma on period 1 contributions, this evidence suggests that charismatic leadership may be particularly helpful for helping groups maintain collective confidence in one another, thereby dampening the decline of contributions with repeated play.

All standard errors reported in Table 4 are adjusted for clustering at the group level. These estimations are therefore based on the assumption that observations across groups within the same session can be regarded as independent. This approach is clean for the first and the second experiment (models 1 through 8) because there is no interaction between participants across groups after entering the laboratory in those settings, and from the beginning to the end of the experiment subjects are alone in their cubicle in the laboratory. In Experiment 3 (models 9 through 12), in contrast, subjects view the video collectively in front of a TV after having read the experimental instructions in their cubicles. As mentioned before, we made sure that people did not communicate when watching the video together. Nevertheless, one might argue that non-verbal forms of interaction might have taken place that render the assumption of independence across groups questionable. We therefore also report the result of a non-parametric rank-sum test using session averages of contributions as independent observations.

The rank-sum test confirms the statistical significance of the positive effect of charisma on contributions in Experiment 3 (p = 0.047).[38]

4.4 Discussion of Combined Laboratory Evidence

The evidence of the effectiveness of charisma across our laboratory experiments is mixed. Experiment 1 finds a modestly positive, but statistically insignificant effect of charisma. Contrary to our predictions Experiment 2 has smaller effects. Finally, Experiment 3 yields the strongest and statistically significant evidence of a positive effect of charisma on contributions.

In this section, we combine the data of all our laboratory studies to further analyze the overall effect of charisma on contributions to the public good. As reported above, the three experiments were conducted in different time periods so that comparisons across experiments are not as clean as comparisons within experiments. However, all experiments were conducted in the same laboratory using the same subject pool and protocols, so we find it meaningful and important to report overall comparisons.

Table 5 reports regressions using data from all three laboratory experiments. We apply a conservative approach and adjust standard errors for clustering at the session level (43 clusters).[39] We first estimate the combined effect of charisma across all experiments, after removing the level effect of the different experiments. For this purpose, we regress contributions on a treatment dummy for the charisma condition and experiment dummies "Charity" (Experiment 2) and "Collective Viewing" (Experiment 3). We report both OLS and Tobit estimations (see Columns (1) and (2)). These regressions reveal that the overall effect of charisma on contributions in all laboratory experiments is positive and at least marginally significant (p = 0.060 (OLS), p = 0.029 (Tobit)).

[Table 5 about here]

Next, we add interaction terms between the charisma dummy and the experiment dummies to partial out the experiment-specific effects of charisma. These specifications yield new insights regarding the impact of the collective viewing experience on contribution behavior. By comparing the two interaction terms ("Charisma x Charity" vs. "Charisma x Collective Viewing") we can perform a simple diff-in-diff analysis of the treatment effects of charisma in Experiments 2 and 3. Such an analysis reveals that the treatment effect in

---

[38] Because of the relatively small number of sessions (12), we do not report regressions with standard errors adjusted for clustering at the session level with data from Experiment 3 alone. However, in section 4.4 we perform such regressions with pooled data from all our laboratory experiments (43 sessions).

[39] We exclude data from the control treatment (no speech) in Experiment 1 (4 sessions), yielding a total of 43 sessions.

Experiment 3 is roughly 4.5 points larger than in Experiment 2. This difference is marginally significant at the 10% level (F-Test: Charisma x Charity = Charisma x Collective Viewing, p = 0.099 (OLS), p = 0.110 (Tobit)). Moreover, the fact that the experiment dummies for "Charity" (Experiment 2) and "Collective Viewing" (Experiment 3) are almost identical suggests that watching the speech together makes no difference when the speech is not very charismatic (F-Test: Charity = Collective Viewing, $p = 0.994$ (OLS), $p = 0.873$ (Tobit)). When the speech is charismatic, in contrast, the group experience seems to matter in that it substantially increases contributions across experiments (F-Test: Charity + Charisma x Charity = Collective Viewing + Charisma x Collective Viewing, p = 0.002 (OLS), p = 0.031 (Tobit)).[40]

5   Conclusion

In the organizational literature, a leader's charismatic properties are often given great importance for motivating follower effort. However, this potential means of influencing economic behavior ranging from worker effort to voluntary pro-social acts has been largely ignored by economic research. To the best of our knowledge, ours is the first paper that presents evidence for a causal link between variation in a leader's charisma and followers' performance in economically relevant settings both with real workers and students.

We report evidence from field and laboratory experiments that investigate whether a leader's charisma—in the form of a stylistically different motivational speech—can induce individuals to undertake personally costly but socially beneficial actions. Our field experiment provides causal evidence for the motivation-enhancing effect of a charismatic speech, but it leaves open the question of how general such a result is and what kinds of mechanisms might underlie its effectiveness. To help address these questions, we present the results of a series of laboratory experiments. The laboratory studies suggest that charisma's effectiveness varies and appears to depend on the social context in which the speech is delivered. In particular, we replicate the positive impact of charisma on prosocial behavior when participants listen to the speech together as a group, but the effects are much smaller when the speech is observed individually. Our combined results suggest that the co-presence of followers when attending to a charismatic speech may be an important factor. Because we do not find evidence that the collective viewing of the speech significantly affects individuals' expectations about others (see Hermalin, 2017 for a theoretical account of such an effect of charisma), we believe that the

---

[40] It is also noteworthy that the regressions confirm that the treatment effect in Laboratory Experiment 3 remains significant when standard errors are adjusted for clustering at the session level (F-Test: Charisma + Charisma x Collective Viewing = 0, p = 0.019 (OLS), p = 0.014 (Tobit)).

most plausible explanation for this result is that charisma nurtures a collective identity (Shamir et al., 1993), which may be particularly valuable in maintaining a group's resilience in the face of a persistent challenge.[41] Given that charisma can be taught (Antonakis et al., 2011), we believe that documenting its effectiveness as a leadership technique is important for the practice of management.

We see our paper as the starting point for a new research agenda. Future research should explore the scope and applicability of our results. Both the field and the laboratory experiments presented in this paper focus on simple, short-term tasks with moral implications. It is thus important to also explore the effects of charisma interventions in a larger variety of settings, involving more varied types of tasks, time horizons and populations of both leaders and followers. Indeed, recent research builds on our work (Meslec et al, 2020), finding that charisma, via video manipulation, can produce higher effort in a cognitive task; this charisma manipulation also works with on-line workers, who were exposed to a short written charismatic text (Fest et al., in press, Study 2). A recent meta-analysis confirms a moderate effect size across a variety of contexts (Ernst et al. (in press)).

Future work should further investigate how variation in characteristics of the task, such as the precise moral nature, might influence the effects of charisma. If workers were stuffing envelopes with advertisement materials for a morally neutral product, would a charismatic speech be effective for inducing effort? It also seems worthwhile to further understand whether charisma effects fade over time and whether new and varied charismatic speeches can sustain a positive impact on effort.

Moreover, one limitation of the experiments in this paper is that we manipulated charisma by simultaneously changing a number of tactics. As a consequence, we cannot determine whether some of the tactics work better than others and how they interact with each other and with the environment. It is possible that the effectiveness of different tactics depends systematically on elements of the environment (e.g. the type of behavior that the leader wants to stimulate, the audience that the speech targets, or the nature of the organization in which followers are situated). In order to provide detailed managerial advice on how to use charisma as a motivational tool, it is important to have a better understanding of these details.

Finally, charisma may be a useful motivational mechanism in settings where standard economic approaches are not helpful. For instance, in complex, multi-task settings in which

---

[41] Please see Appendix I for additional analysis that provides some further empirical support for the interpretation that charisma fosters resilience. The analysis presented in Appendix I is inconsistent with alternative channels such as the idea that charisma fortifies conditional cooperation.

some dimensions of performance are hard to measure, simple pay-for-performance contracts may induce distorted attention and/or unproductive activities (Holmström & Milgrom 1991, Baker 1992). Similarly, it has also been found that output-based pay can be harmful for performance if the worker's task requires creativity (Ariely et al., 2009; Byron & Khazanchi 2012). So far, economists have had difficulties in proposing successful motivation procedures in those situations. Charismatic leadership might present a solution to these kinds of problems.

References:

Alimo-Metcalfe, B., & Alban-Metcalfe, R. J. (2001). The development of a new transformational leadership questionnaire. Journal of Occupational & Organizational Psychology, 79, 1-27.

Angrist, J. D., & Pischke, J. S. (2008). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. Journal of Economic Perspectives, 24(2), 3-30.

Antonakis, J., Bastardoz, N., Jacquart, P. & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. Annual Review of Organizational Psychology and Organizational Behavior, 3(1), 293-319.

Antonakis, J., Bendahan, S., Jacquart, P. & Lalive, R. (2010). On making causal claims: A review and recommendations. The Leadership Quarterly, 21(6), 1086-1120.

Antonakis, J., Fenley, M. & Liechti, S. (2011). Can charisma be taught? Tests of two interventions. The Academy of Management Learning and Education, 10(3), 374-396.

Antonakis, J., Fenley, M. & Liechti, S. (2012). Learning charisma: Transform yourself into someone people want to follow. Harvard Business Review, June, 127-130.

Ariely, D., Gneezy, U., Loewenstein, G. & Mazar, N. (2009). Large stakes and big mistakes. Review of Economic Studies, 76(2), 451-469.

Baker, G. P. (1992). Incentive contracts and performance measurement. Journal of Political economy, 100(3), 598.

Banks, G. C., Engemann, K. N., Williams, C. E., Gooty, J., McCauley, K. D. & Medaugh, M. R. (2017). A meta-analytic review and future research agenda of charismatic leadership. The Leadership Quarterly, 28, 508-529.

Barnard, C. I. (1940). The nature of leadership, Cambridge. MA: The Harvard Universtiy Press.

Bass, B. M. (1985). Leadership and performance beyond expectations. New York: The Free Press.

Bass, B. M., & Avolio, B. J. (1995). MLQ Multifactor leadership questionnaire for research: Permission set. Redwood City, CA: Mindgarden.

Bass, B. M., Waldman, D. A., Avolio, B. J., & Bebb, M. (1987). Transformational leadership and the falling dominoes effect. Group and Organization Studies, 12(1), 73-87.

Bénabou, R., Falk, A., & Tirole, J. (2018). Narratives, imperatives, and moral reasoning. National Bureau of Economic Research (No. w24798).

Bott, K.M., Cappelen, A.W., Sørensen, E.Ø., & Tungodden, B. (2020) You've got mail: A randomized field experiment on tax evasion. Management Science, 66(7), 2801-2819.

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. Management Science, 62(11), 3085-3099.

Brandts, J. & Cooper, D.J. (2007). It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. Journal of the European Economic Association, 5(6), 1223-1268.

Brandts, J., Cooper, D.J. & Weber, R.A. (2015). Legitimacy, communication, and leadership in the turnaround game. Management Science, 61(11), 2627-2645.

Burns, J. M. (1978). Leadership. New York: Harper & Row.

Byron, K. & Khazanchi, S. (2012). Rewards and creative performance: A meta-analytic test of theoretically derived hypotheses. Psychological Bulletin, 138(4), 809.

Conger, J. A., Kanungo, R. N., Menon, S. T., & Mathur, P. (1997). Measuring charisma: dimensionality and validity of the Conger-Kanungo scale of charismatic leadership. Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 14(3), 290-301.

Cooper, W. H. & Richardson, A. J. (1986). Unfair comparisons. Journal of Applied Psychology, 71(2), 179-184.

Cronshaw, S. F. & Lord, R. G. (1987). Effects of categorization, attribution, and encoding processes on leadership perceptions. Journal of Applied Psychology, 72(1), 97-106.

D'Adda, G., Darai, D., Pavanini, N. & Weber, R. 2017. Do leaders affect ethical conduct? An experimental study. Journal of the European Economic Association, 15(6), 1177–1213.

Dal Bó, E. & Dal Bó, P. 2014. "Do the right thing": The effects of moral suasion on cooperation. Journal of Public Economics 117, 28-38.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. Journal of Personality and Social Psychology, 35(1), 1-11.

De Quidt, J., Haushofer, J. & Roth, C. (2018). Measuring and bounding experimenter demand. American Economic Review, 108(11), 3266-3302.

Den Hartog, D. N. & Verburg, R. M. (1997). Charisma and rhetoric: Communicative techniques of international business leaders. The Leadership Quarterly, 8(4), 355-391.

De Hoogh, A., Den Hartog, D., & Koopman, P. (2004). De ontwikkeling van de CLIO: een vragenlijst voor charismatisch leiderschap in organisaties. Gedrag en Organisatie, 17(5), 354-381.

Deutsch, M., Epstein, Y., Canavan, D., Gumpert, P., 1967. Strategies of inducing cooperation: An experimental study. The Journal of Conflict Resolution, 11 (3), 345-360.

Downton, J. V. (1973). Rebel leadership: Commitment and charisma in the revolutionary process. New York: The Free Press.

Eden, D. (2020). The science of leadership: A journey from survey research to field experimentation. The Leadership Quarterly, 101472.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., and Schudy, S. (2019). A field experiment on leadership and team performance in non-routine analytical team tasks. Working Paper.

Ernst, B. A., Banks, G., Loignon, A. C., Frear, K. A., Williams, C. E., Arciniega, L. M., et al. (in press). Virtual Charismatic Leadership and Signaling Theory: A Prospective Meta-Analysis in Five Countries. The Leadership Quarterly.

Fest, S., Kvaløy, O., Nieken, P., & Schöttner, A. (in press). How (not) to motivate online workers: Two controlled field experiments on leadership in the gig economy. The Leadership Quarterly.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2), 171-178.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters, 71(3), 397-404.

Fischbacher, U. & Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. American Economic Review, 100(1), 541-56.

Fischer, T., Hambrick, D. C., Sajons, G. B., & Van Quaquebeke, N. (2020). Beyond the ritualized use of questionnaires: Toward a science of actual behaviors and psychological states. The Leadership Quarterly, 31(4), 101449.

Frese, M., Beimel, S. & Schoenborn, S. (2003). Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision. Personnel Psychology, 56, 671-697.

Gneezy, U., & List, J. A. (2006). Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. Econometrica, 74(5), 1365-1384.

Grant, A. M. (2008). The significance of task significance: job performance effects, relational mechanisms, and boundary conditions. Journal of Applied Psychology, 9, 108–124.

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

Gottfredson, R. K., Wright, S. L., & Heaphy, E. D. (2020). A critique of the Leader-Member Exchange construct: Back to square one. The Leadership Quarterly, 31(6), 101385.

Güntner, A. V., Klonek, F. E., Lehmann-Willenbrock, N., & Kauffeld, S. (2020). Follower behavior renders leader behavior endogenous: The simultaneity problem, estimation challenges, and solutions. The Leadership Quarterly, 101441.

Güth, W., Levati, M. V., Sutter, M. & Van Der Heijden, E. (2007). Leading by example with and without exclusion power in voluntary contribution experiments. Journal of Public Economics, 91(5-6), 1023-1042.

Haushofer, J., & Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. The Quarterly Journal of Economics, 131(4), 1973-2042.

Hermalin B. E. (1998). Toward an economic theory of leadership: Leading by example. American Economic Review, 88, 1188-1206.

Hermalin B. E. (2017). At the helm, Kirk or Spock? The pros and cons of charismatic leadership. Working Paper.

Hogg, M. A. (2001). A social identity theory of leadership. Personality & Social Psychology Review, 5, 184-200.

Holmström B. (1982). Moral hazard in teams. The Bell Journal of Economics, 13, 324-340.

Holmström, B. & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law Economics & Organization, 7, 24.

House, R. J. (1977). A 1976 Theory of Charismatic Leadership. In J. G. Hunt & L. L. Larson (Eds.), The Cutting Edge. Carbondale: Southern Illinois: University Press.

Howell, J. M. & Frost, P. J. (1989). A laboratory study of charismatic leadership. Organizational Behavior and Human Decision Processes, 43(2), 243-269.

Jacquart, P. & Antonakis, J. (2015). When does charisma matter for top-level leaders? Effect of attributional ambiguity. Academy of Management Journal, 58, 1051-1074.

Katz, D., & Kahn, R. L. (1978). The social psychology of organizations. New York: John Wiley & Sons.

Koopman, P. A. R. (1984). Confidence intervals for the ratio of two binomial proportions. Biometrics, 40(2), 513-517.

Kosfeld, M. & Neckermann, S. (2011). Getting more work for nothing? Symbolic awards and worker performance. American Economic Journal: Microeconomics, 3,86–99.

Kosfeld, M., Neckermann, S. & Yang, X. (2014). Knowing That You Matter, Matters! The Interplay of Meaning, Monetary Incentives, and Worker Recognition. IZA Discussion Paper 8055.

Kosfeld, M., & Rustagi, D. (2015). Leader punishment and cooperation in groups: Experimental field evidence from commons management in Ethiopia. American Economic Review, 105(2), 747-83.

Kube, S., Maréchal, M. A., & Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. American Economic Review, 102(4), 1644-62.

Kvaløy, O., Nieken, P. & Schöttner, A. (2015). Hidden benefits of reward: A field experiment on motivation and monetary incentives. European Economic Review, 76, 188-199.

Kvaløy, O. & Schöttner, A. (2015). Incentives to motivate. Journal of Economic Behavior & Organization, 116, 26-42.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33(1), 159-174.

Lazear, E. (2000). Performance pay and productivity. American Economic Review, 90(5): 1346-1361.

Lazear, E. P. & Rosen, S. (1981). Rank and order tournament: An optimal labor contract. Journal of Political Economy, 89. 841-864.

Lonati, S., Quiroga, B. F., Zehnder, C. & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. Journal of Operations Management, 64, 19-40.

McKenzie, D., & Ozier, O. (2019). Why ex-post power using estimated effect sizes is bad, but an ex-post MDE is not. World Bank Development Impact Blog.

Meslec, N., Curseu, P., Fodor, O. C., & Kenda, R. (2020). Effects of charismatic leadership and rewards on individual performance. The Leadership Quarterly, 31(6): 101423.

Minozzi, W., Neblo, M.A., Esterling K.M. & Lazer, D.M.J. 2015. Field experiment evidence of substantive, attributional, and behavioral persuasion by members of Congress in online town halls. Proceedings of the National Academy of Sciences, 112(13): 3937–3942.

Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on follower's trust in leader, satisfaction, and organizational citizenship behaviors. The Leadership Quarterly, 1(2), 107-142.

Potters, J., Sefton, M. & Vesterlund, L. (2007). Leading-by-example and signaling in voluntary contribution games: an experimental study. Economic Theory, 33, 169-182.

Sajons, G. B. (2020). Estimating the causal effect of measured endogenous variables: A tutorial on experimentally randomized instrumental variables. The Leadership Quarterly, 31(5), 101348.

Shamir, B., House, R. J. & Arthur, M. B. (1993). The motivational effects of charismatic leadership: A self-concept based theory. Organization Science, 4(4), 577-594.

Shamir, B., Arthur, M. B. & House, R. J. (1994). The rhetoric of charismatic leadership: A theoretical extension, a case study, and implications for research. The Leadership Quarterly, 5(1), 25-42.

Shea, C. M. & Howell, J. M. (1999). Charismatic leadership and task feedback: A laboratory study of their effects on self-efficacy and task performance. The Leadership Quarterly, 10, 375-396.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. Review of Economic Studies, 71, 513–534.

Sieweke, J., & Santoni, S. (2020). Natural experiments in leadership research: An introduction, review, and guidelines. The Leadership Quarterly, 31(1), 101338.

Towler, A. J. (2003). Effects of charismatic influence training on attitudes, behavior, and performance. Personnel psychology, 56(2), 363 - 381.

Tur, B., Harstad, J., & Antonakis, J. (2021). Effect of charismatic signaling in social media settings: Evidence from TED and Twitter. The Leadership Quarterly.

Turner, J. C., Oakes, P. J., Haslam, S. A. & McGarty, C. (1994). Self and collective: Cognition and social context. Personality and social psychology bulletin, 20(5), 454-463.

van Knippenberg, D., & Sitkin, S. B. (2013). A critical assessment of charismatic-transformational leadership research: Back to the drawing board? The Academy of Management Annals, 7(1), 1-60.

Van Vugt, M. & De Cremer, D. (1999). Leadership in social dilemmas: The effects of group identification on collective actions to provide public goods. Journal of Personality and Social Psychology, 76(4), 587.

Waldman, D. A., Bass, B. M., & Yammarino, F. J. (1990). Adding to contingent-reward behavior-The augmenting effect of charismatic leadership. Group & Organization Studies, 15(4), 381-394.

Weber, M. (1947). The theory of social and economic organization (T. Parsons, Trans.). New York: The Free Press.

Willner, A. R. (1984). The spellbinders: Charismatic political leadership. New Haven: CT.: Yale University Press.

Yukl, G. A. (1999). An evaluation of conceptual weaknesses in transformational and charismatic leadership theories. The Leadership Quarterly, 10(2), 285-305.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. Experimental Economics, 13(1), 75-98.

Table 1: Descriptive Statistics, Field Experiment

Treatment	Baseline	Piece Rate	Charisma	p-value
Workers' characteristics				
Average age of workers (in years)	32.2	32.6	32.5	0.99
Percentage of female workers	0.46	0.57	0.61	0.41
Experimental behavior				
Average earnings of workers (in GBP)	28.4	37.2	28.4	
Av. number of letters completed	230.94	277.70	271.22	
Standard deviation	*86.66*	*79.99*	*80.76*	
Max. number of letters completed	340	349	350	
Min. number of letters completed	90	60	100	
25th percentile	*150*	*250*	*200*	
Median number of letters completed	210	317	320	
75th percentile	*320*	*330*	*337*	
Av. quality index	0.86	0.88	0.90	
Rel. freq. of perfect quality	0.21	0.30	0.32	
Av. cost per letter completed (in GBP)	0.123	0.134	0.105	
Av. time spent on the task (in minutes)	245	267	263	
Standard deviation	*64.57*	*68.40*	*97.78*	
Max. time spent on the task	370	480	480	
Min. time spent on the task	150	160	60	
Median time spent on the task	240	270	240	
Rel. freq. of help from friends or family	0.12	0.19	0.18	

Notes: The table provides descriptive statistics for relevant variables of our field experiment. Balance is tested by running a regression of workers' characteristics on treatment dummies (Piece-rate, Charisma): we report p-values from a Wald test of beta(Piece-rate) = beta(Charisma) = 0 in Column 4. In the baseline and the charisma treatments the workers' wage is fixed to GBP 28.4 (this corresponds to 4.5 times the hourly minimum wage). In the piece rate condition workers who completed more than 220 envelopes received GBP 0.12 for each additional envelope. Workers had received the materials to complete a maximum of 350 letters. The quality index contains 20 items (see section 3.4 for details). Perfect quality is unity if all checked items of the quality index have been passed and zero otherwise. The average cost per letter has been calculated at the treatment level (i.e., the number reported corresponds to the sum of worker earnings divided by the sum of letters completed by all workers in a treatment). One worker (in the charisma treatment) did not respond to the age question so that we have one missing observation. The time spent on the task and the availability of help from friends or family are self-reported by workers and could therefore not be verified. Some workers have not responded to these questions so that we have 8 missing observations for each variable (time: 3 Baseline, 4 Piece Rate, 1 Charisma / help: 2 Baseline, 4 Piece Rate, 2 Charisma).

Table 2: Regression Analysis of Treatment Effects in the Field Experiment

Dependent variable	Letters completed (1)	Letters completed (2)	Cost per letter (3)	Quality index (4)	Perfect quality (5)
Panel A: no clustering					
Piece rate	46.757**	47.633*	0.006	0.023	0.094
	(20.656)	(24.860)	(0.018)	(0.034)	(0.110)
Charisma	40.277**	58.509***	-0.028**	0.041	0.119
	(19.346)	(21.123)	(0.014)	(0.030)	(0.103)
Constant	230.943***	212.711***	0.146***	0.857***	0.206***
	(14.647)	(16.931)	(0.011)	(0.025)	(0.070)
Panel B: clustering by session					
Piece rate	46.757**	47.633*	0.006	0.023	0.094
	(22.484)	(27.946)	(0.017)	(0.031)	(0.113)
Charisma	40.277*	58.509**	-0.028*	0.041	0.119
	(20.797)	(26.228)	(0.015)	(0.029)	(0.105)
Constant	230.943***	212.711***	0.146***	0.857***	0.206***
	(15.275)	(18.895)	(0.011)	(0.021)	(0.078)
Missing equal to "No work"?	No	Yes	No	No	No
R-Squared	0.059	0.070	0.051	0.019	0.013
Observations	106	111	106	104	104

Notes: In columns (1) and (2) the dependent variable is the number of letters completed at the worker level. In column (3) the dependent variable is the cost per letter measured at the worker level. The dependent variable in column (4) is an index that measures the fraction of passed items in a set of 20 quality checks at the worker level. The dependent variable in column (5) is unity if all 20 quality checks have been passed and zero otherwise. The variables "Piece Rate" and "Charisma" are indicator variables for treatments; the omitted category is "Baseline". Column (1) considers only workers who returned the completed envelopes, thus setting to missing observations corresponding to workers who dropped out of the experiment after the treatment, but before performing the work task. Column (2) instead includes these observations, setting the output of workers who dropped out of the experiment equal to zero. The constant in column (3) does not correspond to the average cost per letter for the baseline treatment reported in Table 1. The reason is that in Table 1 the cost per letter has been calculated at the treatment level (see the notes to Table 1 for more details). Because of a coordination problem we failed to collect the quality information for two workers, which explains why we have only 104 observations in columns (4) and (5). Panel A reports robust standard errors, while Panel B displays robust standard errors clustered at the session level in parentheses, with Moulton method used to correct for the small number of clusters; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3: Regression Analysis of Treatment Effects in the Field Experiment (with covariates)

Dependent variable	Letters completed (1)	Cost per letter (2)	Time spent on task (3)	Time spent on task (4)	Help from friends/fam. (5)	Help from friends/fam. (6)
Panel A: no clustering						
Piece rate	45.373**	0.007	21.611	23.741	0.071	0.085
	(21.257)	(0.020)	(17.569)	(18.219)	(0.097)	(0.097)
Charisma	36.836*	-0.025*	17.788	13.349	0.058	0.079
	(20.165)	(0.015)	(19.251)	(19.192)	(0.085)	(0.086)
Female	9.490	-0.008		-2.739		-0.107
	(17.129)	(0.014)		(17.234)		(0.084)
Age	0.796	-0.001*		1.385**		-0.000
	(0.567)	(0.000)		(0.628)		(0.004)
Constant	200.977***	0.171***	245.313***	202.137***	0.121**	0.185
	(24.421)	(0.018)	(11.411)	(21.855)	(0.058)	(0.158)
Panel B: clustering by session						
Piece rate	45.373**	0.007	21.611	23.741	0.071	0.085
	(22.495)	(0.017)	(21.313)	(20.614)	(0.098)	(0.099)
Charisma	36.836*	-0.025	17.788	13.349	0.058	0.079
	(21.070)	(0.015)	(19.145)	(18.652)	(0.088)	(0.090)
Female	9.490	-0.008		-2.739		-0.107
	(18.335)	(0.013)		(16.311)		(0.080)
Age	0.796	-0.001		1.385**		-0.000
	(0.697)	(0.001)		(0.642)		(0.003)
Constant	200.977***	0.171***	245.313***	202.137***	0.121*	0.185
	(29.851)	(0.022)	(14.270)	(26.968)	(0.065)	(0.130)
Missing equal to "No work"	No	No	No	No	No	No
R-squared	0.073	0.064	0.013	0.063	0.007	0.027
Observations	105	105	98	97	98	97

Notes: In column (1) the dependent variable is the number of letters completed at the worker level. In column (2) the dependent variable is the cost per letter measured at the worker level. The variables "Piece Rate" and "Charisma" are indicator variables for treatments; the omitted category is "Baseline". These regressions also control for age and gender of workers. We lose one observation, because one of the workers did not report her age in the questionnaire. The dependent variable in columns (3) and (4) is the time spent on the task self-reported by the worker. The dependent variable in columns (5) and (6) is an indicator variable of whether the worker had help from friends or family. For both variables we report uncontrolled treatment effects and estimations controlling for age and gender. We lose another 8 observations in these regressions, because some workers did not fill in the section of the questionnaire that contained the questions on the work task (see also the notes of Table 1 for more details). Panel A reports robust standard errors, while Panel B displays robust standard errors clustered at the session level in parentheses, with Moulton method used to correct for the small number of clusters; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Regression Analysis, Laboratory Experiments

Experiment	Laboratory Experiment 1				Laboratory Experiment 2				Laboratory Experiment 3			
Estimator	OLS	OLS	Tobit	Tobit	OLS	OLS	Tobit	Tobit	OLS	OLS	Tobit	Tobit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Panel A											
Dependent variable	Contribution											
Charisma	1.097	1.132	1.622	2.073	0.320	0.643	1.309	1.744	2.515**	1.385	5.694**	4.992**
	(1.178)	(1.118)	(1.865)	(1.997)	(1.111)	(0.856)	(2.386)	(2.147)	(1.154)	(0.905)	(2.511)	(2.416)
Control	-2.188	-3.369**	-3.409	-5.349**								
	(1.488)	(1.451)	(2.133)	(2.236)								
Period		-0.996***		-1.312***		-0.459***		-0.720***		-0.585***		-0.872***
		(0.069)		(0.095)		(0.088)		(0.141)		(0.101)		(0.151)
Charisma x Period		-0.006		-0.081		-0.059		-0.081		0.205		0.116
		(0.119)		(0.166)		(0.135)		(0.230)		(0.147)		(0.269)
Control x Period		0.215		0.357**								
		(0.134)		(0.174)								
Constant	10.057***	15.537***	12.638***	19.769***	13.590***	16.115***	19.047***	22.931***	13.582***	16.797***	18.139***	22.865***
	(0.891)	(0.854)	(1.479)	(1.614)	(0.735)	(0.542)	(1.776)	(1.600)	(0.968)	(0.647)	(2.150)	(1.977)
Observations	4,360	4,360	4,360	4,360	3,000	3,000	3,000	3,000	2,280	2,280	2,280	2,280
R-squared	0.019	0.124			0.000	0.033			0.031	0.072		
	Panel B											
Dependent variable	Belief											
Charisma	1.120	1.269	1.387	1.741	0.476	1.087*	1.344	1.968*	2.081*	0.913	2.867*	1.658
	(1.095)	(0.782)	(1.497)	(1.192)	(1.010)	(0.652)	(1.553)	(1.140)	(1.043)	(0.707)	(1.628)	(1.147)
Control	-2.467*	-3.343***	-3.266*	-4.583***								
	(1.371)	(0.932)	(1.706)	(1.211)								
Period		-0.814***		-0.917***		-0.250***		-0.268**		-0.350***		-0.353**
		(0.096)		(0.114)		(0.083)		(0.108)		(0.114)		(0.142)
Charisma x Period		-0.027		-0.064		-0.111		-0.116		0.212		0.218
		(0.147)		(0.177)		(0.134)		(0.191)		(0.145)		(0.195)
Control x Period		0.159		0.245								
		(0.156)		(0.182)								
Constant	11.012***	15.490***	12.321***	17.297***	13.779***	15.154***	15.406***	16.857***	13.743***	15.669***	15.235***	17.154***
	(0.822)	(0.538)	(1.139)	(0.863)	(0.671)	(0.457)	(1.084)	(0.800)	(0.839)	(0.489)	(1.254)	(0.871)
Observations	4,360	4,360	4,360	4,360	3,000	3,000	3,000	3,000	2,280	2,280	2,280	2,280
R-squared	0.032	0.133			0.002	0.024			0.036	0.056		

Notes: In Panel A the dependent variable is the contribution to the public good. The variables "Charisma" and "Control" are indicator variables for treatments; the omitted category is "No Charisma". There are four estimations for each of our three laboratory experiments. We estimate pure treatment effects and a specification that controls for treatment-specific time trends. We report OLS and Tobit estimates (to account for the fact that contributions are censored at the maximum level of 20 points). In Panel B the dependent variable is the belief about others' contributions. Standard errors are adjusted for clustering at the group level; *** $p<0.01$, ** $p<0.05$, *$p<0.10$.

Table 5: Regression Analysis (All Laboratory Experiments)

Estimator	OLS	Tobit	OLS	Tobit
	(1)	(2)	(3)	(4)
Dependent variable		Contribution		
Charisma	1.201*	2.508**	1.097	1.687
	(0.620)	(1.145)	(1.185)	(1.936)
Charity	3.105***	5.280***	3.533***	5.546***
	(0.702)	(1.222)	(1.172)	(2.082)
Collective Viewing	4.200***	7.066***	3.525***	5.218**
	(0.812)	(1.539)	(1.238)	(2.087)
Charisma x Charity			-0.777	-0.469
			(1.427)	(2.532)
Charisma x Collective Viewing			1.417	4.057
			(1.572)	(3.049)
Constant	10.004***	12.592***	10.057***	13.002***
	(0.703)	(1.288)	(0.912)	(1.613)
Observations	8,840	8,840	8,840	8,840
R-squared	0.054		0.056	

Notes: The dependent variable is the contribution to the public good. The variable "Charisma" is an indicator variable for the charisma treatment; the omitted category is "No Charisma". The variables "Charity" and "Collective Viewing" are indicator variables for observations from Laboratory Experiments 2 and 3, the omitted category is Laboratory Experiment 1. We exclude data from the control treatment (no speech) in Laboratory Experiment 1. Columns (1) and (2) present OLS and Tobit estimates of the overall treatment effect of charisma across all three experiments (controlling for experiment-specific level effects). In columns (3) and (4) we add interaction effects between charisma and experiments 2 and 3. Standard errors are clustered at the session level; *** $p<0.01$, ** $p<0.05$, *$p<0.10$.

Figure 1: Job Ad used in the Recruitment Process

Notes: This figure displays the original job ad as shown on the agency's webpage.

Figure 2: Model Predictions – Effects of Piece Rate and Charisma
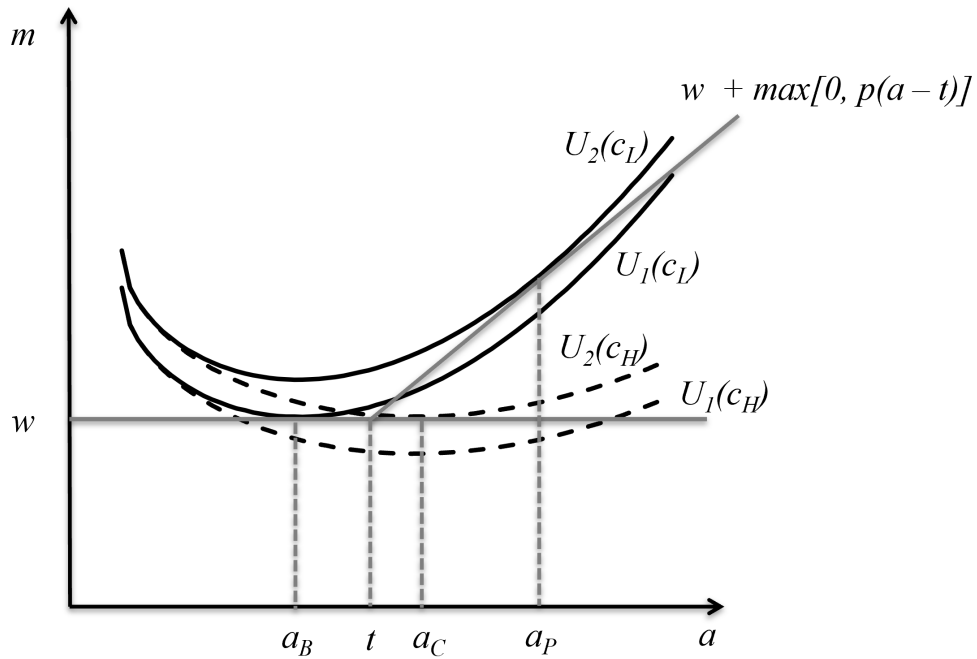
Notes: The figure shows examples of indifference curves in the worker's benefit-output space for both low charisma (c_L, solid black lines) and high charisma (c_H, dashed black lines) of the leader. In the baseline condition and the charisma condition the material benefit corresponds to the fixed wage w, so that the set of reachable points consists of a flat line (horizontal grey line). In the piece rate condition the set of reachable points includes a kink (increasing grey line), because the wage increases if the worker surpasses a threshold t ($m = w + \max[0, p(a - t)]$, where p is the piece rate. The figure displays a utility-maximizing worker's optimal choices for each of the three treatments (the worker always aims at reaching the highest indifference curve possible given the set of reachable points). In the baseline condition the utility maximizing output choice is a_B. The corresponding levels in the charisma condition and the piece rate condition are a_C and a_P.

Figure 3: Number of Letters Stuffed and Cost per Letter across Treatments

Panel A: Average Number of Letters Stuffed
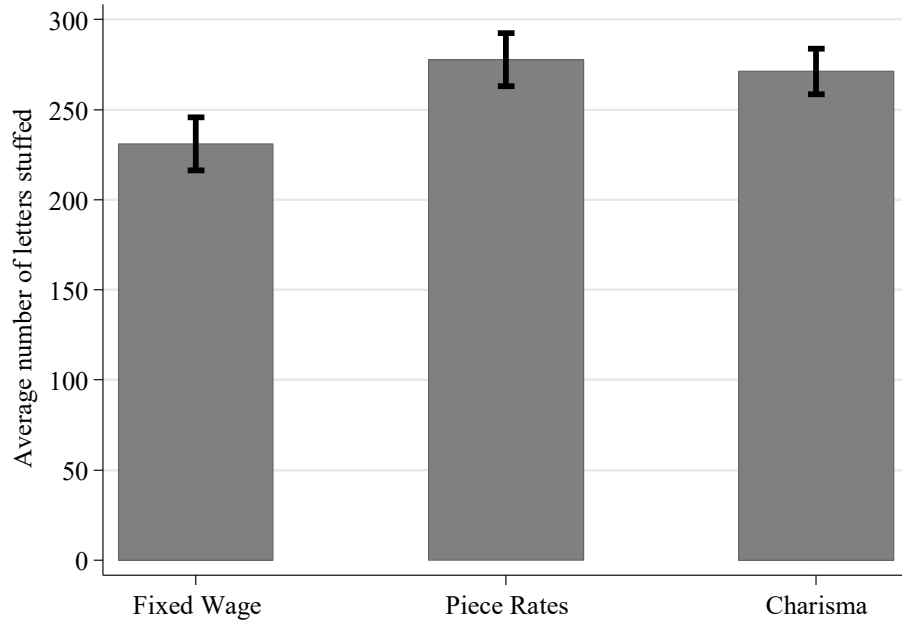
Panel B: Average Cost per Letter
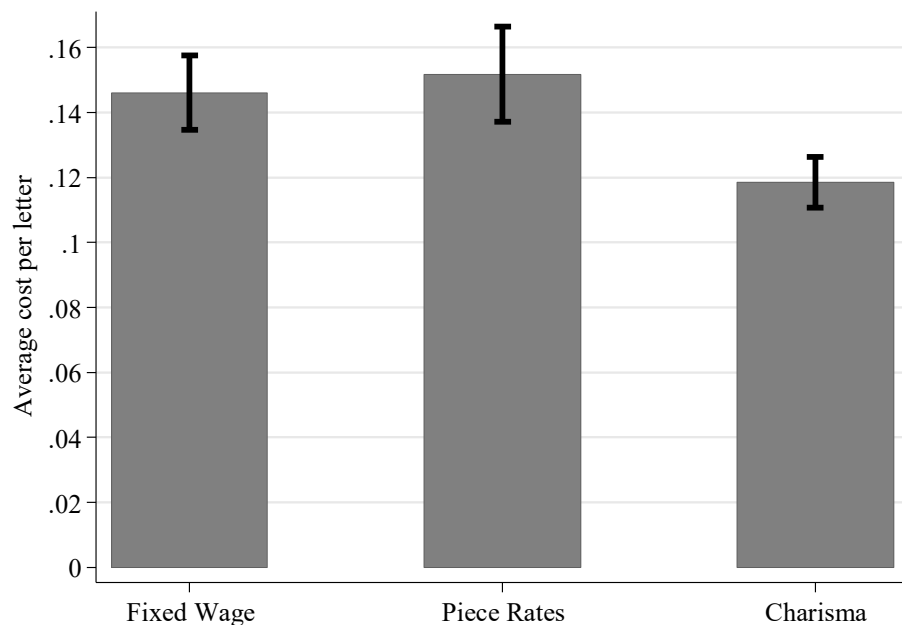
Notes: The figures displays the average number of letters completed (Panel A) as well as the average cost per letter (Panel B) across treatments. The capped bars represent the standard error of the mean. The average cost per letter has been calculated at the worker level before averaging across workers. The displayed levels therefore correspond to the values reported in Table 2 rather than the values reported in Table 1.

Figure 4: Distribution of Number of Letters Stuffed across Treatments
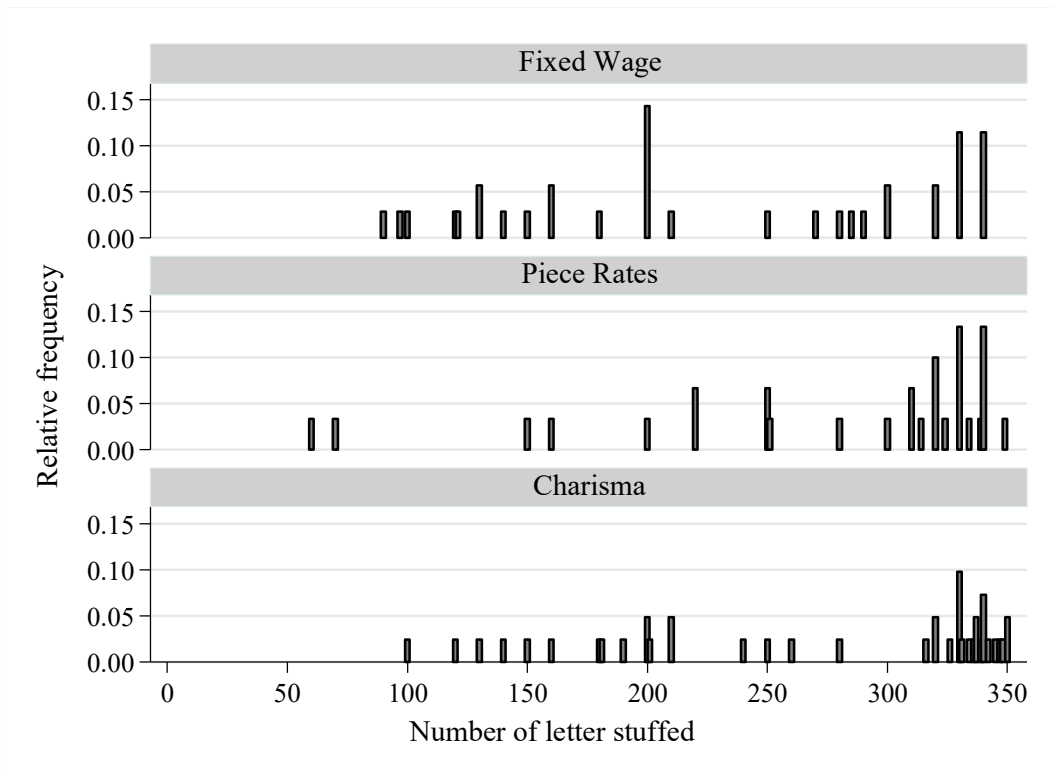
Notes: The figures show the distributions of the number of completed letters across our three treatments.

Figure 5: Model Predictions – Effects of Charisma on Contributions to Public Good
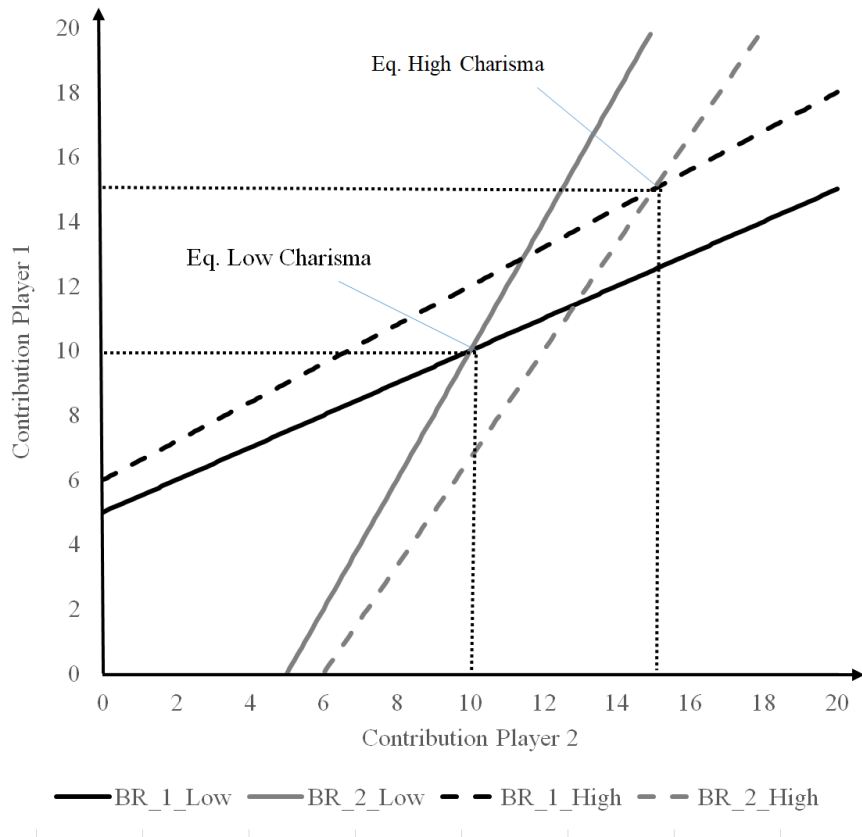
Notes: The figure illustrates how an increase in leader charisma shifts the best response functions of Players 1 (solid and dashed black lines) and 2 (solid and dashed grey lines) and how this increases the equilibrium contributions of the two players. The figure is drawn for the following parameters: $\alpha = 0.8$, $\gamma = 2$, $\delta = 0.2$. Perceived charisma is assumed to increase from $c = 0.5$ to $c = 0.6$.

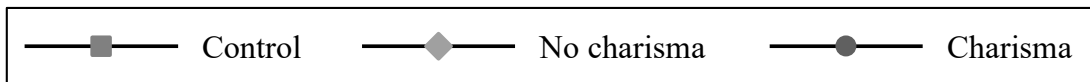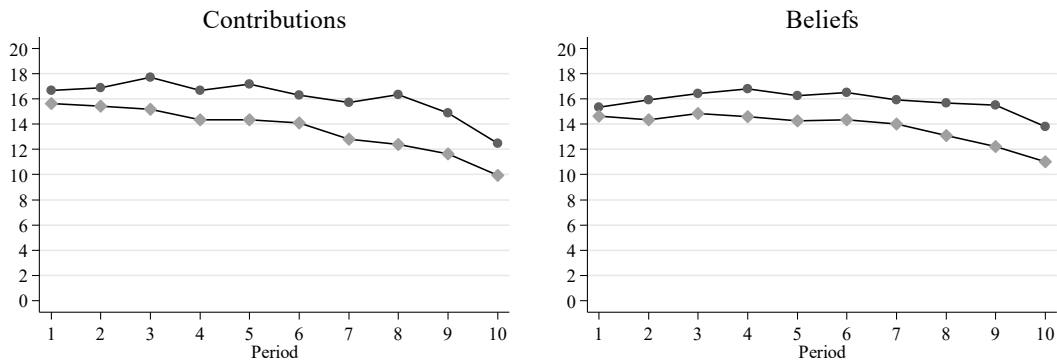# Figure 6: Contributions and Beliefs over Time (Laboratory Experiments)

Panel A: Experiment 1
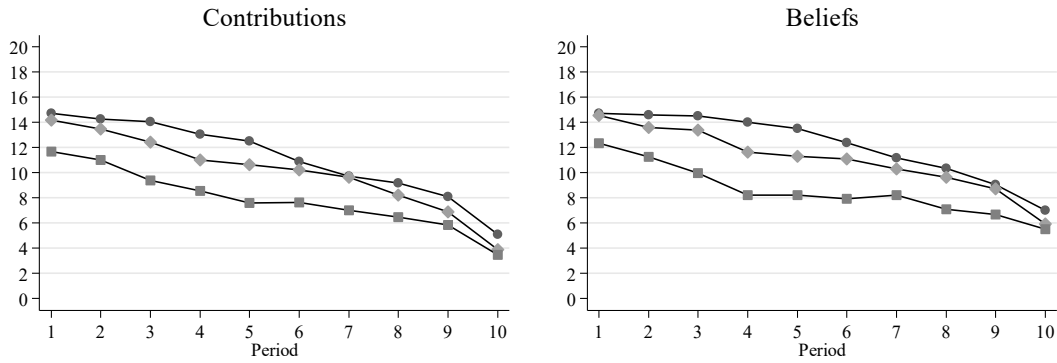
Panel B: Experiment 2
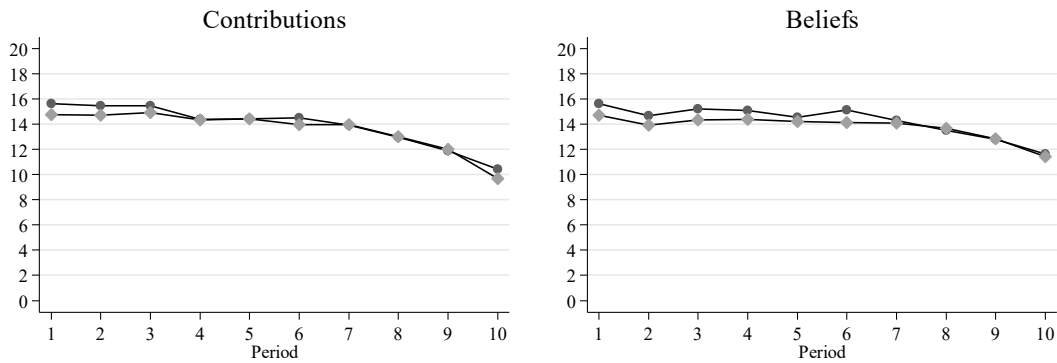
Panel C: Experiment 3

Notes: The figures show the development of public good contributions and beliefs across time