Mutual Knowledge of Social Norms and Political Behavior*

Anselm Hager ⓡ Elnura Kazakbaeva ⓡ Lukas Hensel ⓡ Damir Esenaliev

February 24, 2025

Abstract

Social norms are crucial drivers of human behavior. However, misperceptions of others' opinions may sustain norms and conforming behavior even if a majority opposes the norm. Privately shifting individuals' beliefs about true societal support may be insufficient to change behavior if others are perceived to continue to hold incorrect beliefs ("lack of mutual knowledge"). We conduct a field experiment with 5,201 women in Kyrgyzstan to test whether creating mutual knowledge about social norms affects how perceived social norms influence behavior. We show that providing information about societal support for female political activism alone does not affect women's political engagement. However, when perceived mutual knowledge is created, the effect of information about social norms increases significantly. Using vignette experiments, we show that the effect of mutual knowledge on social punishment is a plausible mechanism behind the behavioral impact. These findings suggest that higher-order beliefs about social norms are an important force linking social norms and behavior.

Keywords: Social norms; higher-order beliefs; political activism; field experiment

1   Introduction

Social norms are powerful predictors of human behavior (Bicchieri, 2005; Fehr and Schurtenberger, 2018). Norms are particularly relevant for women, whose political, economic, and social behavior is subject to social norms (Bursztyn et al., 2017; Cheema et al., 2023; Jayachandran, 2021). However, social norms are dynamic and can rapidly change (Andreoni et al., 2021; Bicchieri, 2016). In particular, information about others' private opinions can lead to marked changes in perceptions of norms and underlying behavior. Famously, Bursztyn et al. (2020a) inform Saudi Arabian men that other men support female labor force participation which leads men to support women in the labor market.

We argue that only receiving information about others' *private* opinions can be insufficient to change in behavior if others hold incorrect beliefs and enforce a social norm despite being privately opposed to it. Take the following example: a woman learns that other community members support female political activism. However, if she still believes that *others* think that society does not support women in politics, she might still be deterred from political action. Her perceptions of others' beliefs may lead to the anticipation that community members will enforce negative norms against their private convictions, for example, because the enforcement action itself is subject to the social norm. The only solution to this problem is making the true support for the social norm *mutual knowledge*. In other words, once a woman learns that i) society supports female engagement and that ii) other women, too, know about this, she should become more likely to engage in politics.

In this paper, we conduct a field experiment with 5,201 women in Kyrgyzstan to test whether mutual knowledge affects how social norms shape female political participation. Our experiment has two main components. First, we randomize which of the two expert estimates of social support for female political participation respondents receive. Specifically, women are truthfully told that an expert thinks that either 1 out of 10 (low group) or 7 out of 10 (high group) people in Kyrgyzstan want women to be more politically active. We use this

active control design that allows the recovery of causal effects independent of pre-treatment beliefs (Haaland et al., 2023). We call this experiment the *social norm treatment* because it mimics typical information provision experiments designed to shift perceptions of social norms.

Second, we create exogenous variation in mutual knowledge by cross-randomizing whether women are told that the information is also given to other community members, the *mutual knowledge treatment*. Technically, we experimentally shift third-order beliefs in addition to shifting second-order beliefs about social norms around female political participation. We then test whether the mutual knowledge treatment increases the effect of the social norm treatment on political participation. Specifically, we measure the impact on women's actual participation in key community meetings as an unobtrusive behavioral outcome for political engagement in local communities.

The experiment yields three main results. First, we show that simply informing respondents about private opinions of others only affects beliefs, but not behavior. Women who receive the high expert opinion update their second-order beliefs by the equivalent of 5.1 percentage points (pp) more positively than women who receive the low expert opinion ($p < 0.001$). However, this shift in second-order beliefs does not affect on women's level of political engagement. Without mutual knowledge, the high group is no more likely to attend the community meeting than the low group.

Second, the mutual knowledge treatment significantly increases the effect of the social norm treatment on behavior. When women receive the mutual knowledge treatment, the gap between the high and low groups increases by 3pp ($p = 0.098$). The behavioral effect is mostly driven by a negative effect of mutual knowledge in the low information group (-3.6pp, $p = 0.006$), whereas there is no significant effect on behavior in the high information group. This result suggests that the effect of mutual knowledge depends on whether the social norm is positive or negative.

Third, we explore mechanisms with additional vignette experiments with 3,760 women in the same communities. The vignettes provide evidence that social punishment and praise are plausible mechanisms behind the observed effect of mutual knowledge. We ask women whether they would react by punishing or praising when they see a woman attend a political meeting. We randomize the scenario in two ways: whether women themselves are told to imagine supporting or opposing female political activism and whether they live in a village with perceived supportive or opposing norms. We find that, when women are told to imagine opposing female activism, they are significantly more likely to punish activism when the village is opposed compared to being in the supportive village scenario. In line with our main treatment effects, we find that alignment does not make a difference when the personal opinion is supportive. We also consider persuasion effects, that is whether mutual knowledge might induce women to change their own opinion about social norms. While we observe mutual knowledge leads to some changes in private opinions, this effect mediates at most 15% percent of the mutual knowledge effect. Overall, we conclude that changes in anticipated social reactions are the most likely channel behind the mutual knowledge effect.

The key contribution of our paper is to demonstrate that creating perceived mutual knowledge about social norms increases the effect of social norm interventions. The existing literature has studied the effect of providing information about social norms on behavior, but without experimentally testing for the role of higher-order beliefs. Typical experimental studies only attempt to shift second-order beliefs about prescriptive social norms (Bursztyn et al., 2020a).[1] An exception is Bursztyn et al. (2020b) who study the effect of public signals about social norms that generate common knowledge on political expression and behavior. However, they do not explicitly test how the public nature of their signal affects its effectiveness. Explicitly documenting the role of higher-order beliefs is important to

---

[1]There is also a large literature on the impact of second-order beliefs about descriptive norms, that is the anticipated behavior of others, on political (Cantoni et al., 2019; Gerber et al., 2008; Hager et al., 2022, 2023; Perez-truglia and Cruces, 2017), consumer (Allcott, 2011; Brandon et al., 2019), compliance (Hallsworth et al., 2017), and prosocial (Chen et al., 2010; Frey and Meier, 2004; Shang and Croson, 2009) behavior. None of these papers study higher-order beliefs. It is also unclear how third-order beliefs about descriptive norms should theoretically affect behavior conditional on second-order beliefs.

understand how social norms shape behavior and how to best design social norm-based interventions.

Second, we relate to a small literature documenting the empirical importance of higher-order beliefs for political behavior.[2] George et al. (2019) experimentally create mutual knowledge about candidate characteristics in an Indian election and document impacts on private vote choices. Similarly, Gottlieb (2016a) experimentally generates common knowledge by providing information about candidates in public settings in Mali and studies the impact on vignette vote choices. Finally, Bursztyn et al. (2023) show that individuals anticipate a reduction in the stigma of expressing dissent when they have a reasonable social cover, leading to more expressions of dissent. Similar to these studies, we find that high-order beliefs affect political behavior. However, we study how higher-order beliefs shape the effect of social norms on behavior. We further provide evidence that public enforcement may drive the effect of norms on publicly observable political actions.

Third, we contribute to the literature studying the role of social norms for female political participation in low- and middle-income countries (Milazzo and Goldstein, 2019; Paxton et al., 2007).[3] For example, Cheema et al. (2023) find that canvassing increases female turnout in Pakistan by more when husbands are also present during the visit. More generally, Gottlieb (2016b) argues that informational resources might not be sufficient to increase participation when social norms remain opposed. We confirm the role of social norms but also show that higher-order beliefs about social norms can be a (necessary) amplifier of their effect on female

---

[2]We also relate to a literature on higher-order beliefs in other fields. Theoretical research has long predicted an important role for higher-order beliefs. Applications include behavioral game theory (Camerer et al., 2004; Crawford and Iriberri, 2007; Crawford et al., 2013), macroeconomics (Angeletos and Jennifer, 2009; Huo and Takayama, 2024; Nimark, 2008), and finance (Banerjee and Kremer, 2010; Cespa and Vives, 2015). Experimental work in various contexts has also demonstrated the importance of higher-order beliefs. For example, Coibion et al. (2021) experimentally shows that higher-order inflation beliefs affect firms' price setting. Similarly, Gorodnichenko and Yin (2024) shift higher-order beliefs about other investors and find an impact on asset holdings. There is also a large number of lab-experimental studies demonstrating that strategic behavior is affected by higher-order beliefs (Costa-Gomes et al., 2001; Dechenaux et al., 2015; Nagel, 1995).

[3]A related literature links social networks and information spillovers to female political participation (Giné and Mansuri, 2018; Prillaman, 2021).
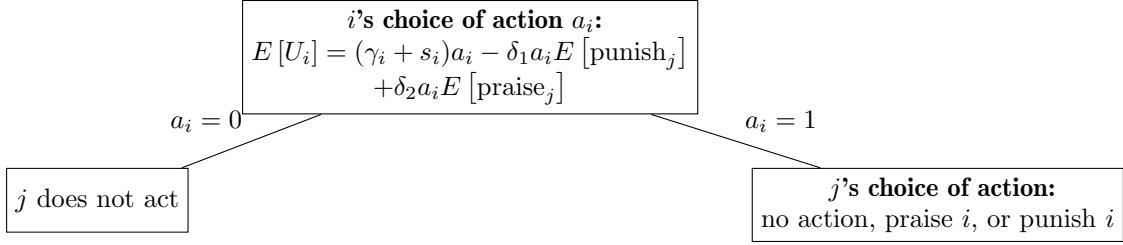
political participation.

Moreover, we provide evidence about the underlying mechanisms that link social norms to changes in political behavior. Theoretical channels for the positive effect of information experiments regarding social norms on behavior point to changed perceptions about potential punishment (Bernhard et al., 2006) or community praise (Bénabou and Tirole, 2006). Our vignette experiments confirm a likely role for social reactions; they show that reactions are subject to social norms themselves, which can explain how higher-order beliefs about social norms influence behavior. This finding also contributes to a large literature documenting punishment behavior in the lab (Abbink et al., 2017; Fehr and Fischbacher, 2004) and in the field (Ensminger and Henrich, 2014). These papers typically explore third-party punishment in response to breaking social norms but neglect the fact that sanctioning behavior itself might be subject to social sanctions. We demonstrate that (anticipated) social sanctions may link higher-order beliefs to behavior.

The paper proceeds as follows. In Section 2 we describe a theoretical framework that motivates the experiment. Section 3 describes the context and experimental design. Sections 4 and 5 describe the main experimental results. Section 6 provides vignette evidence on mechanisms. Section 7 concludes.

2   Conceptual framework

We model the decision to be politically active in a sequential game between N (female) community members indexed by i and the community at large j. Figure 1 displays the structure of the decision problem. In the first stage, i decides whether to engage in a political action a_i. In the second stage, the community j decides whether to react by publicly praising or punishing if i acted during the first stage. We assume that j only reacts to action a_i if their private opinion s_i about the desirability of the political action and their perceptions of prevailing social norms in the community align. The role of perceived social norms is

6

Figure 1: Two period game of political action and social norm enforcement

i's choice of action a_i:
$$E[U_i] = (\gamma_i + s_i)a_i - \delta_1 a_i E\left[\text{punish}_j\right]$$
$$+\delta_2 a_i E\left[\text{praise}_j\right]$$

$a_i = 0$ $a_i = 1$

j does not act

j's choice of action:
no action, praise i, or punish i

Notes: **Figure 1 illustrates the two-period game linking perceived social norms and political activism a_i.**

motivated by the fact that punishment and praise are mostly perceived as publicly observable (see the discussion in Section 6) and, thus, potentially subject to social scrutiny. i anticipates the second-stage reactions and incorporates them into her first-stage decisions, linking third-order beliefs to decisions about political activism. We solve the decision problem by backward induction. We first describe j's decision in period two, before analyzing i's discussion in the first period. We provide a more detailed discussion of the setup and all proofs in Appendix Section C.

Period 2 decision   The community j chooses one of three possibilities: praise for a_i, punish a_i, or do nothing. Their action depends on their private opinion s_i about whether the action a_i is desirable. It also depends on the prevailing social norm (s_-). s_j and s_- can either be supportive of a_i in which case they take a value of 1 or oppose a_i in which case they take a value of zero. j might misperceive others' private opinions, and by implication the social norm s_-, and form beliefs $\widehat{s_-}$ about the prevailing social norm. We assume that j only reacts if private opinion and the (perceived) social norm align to simplify the setup.[4]

---

[4]Making the community choice a stochastic function of perceived social norms complicates the analysis without yielding substantially different insights. Similarly, allowing for the punishment or praise of inaction merely strengthens the existing mechanism without yielding additional theoretical insights.

Period 1 decision for i A woman i decides whether to engage in action a_i according to the following expected utility function:

$$E\left[U_w(a_i, s_i, \tilde{s}_j, \widehat{\tilde{s}_-})\right] = a_i \left(\gamma_i + s_i - \delta_1 E\left[punish_j | \tilde{s}_j, \widehat{\tilde{s}_-}\right] + \delta_2 E\left[praise_j | \tilde{s}_j, \widehat{\tilde{s}_-}\right]\right) \quad (1)$$

where γ_i is the private utility i gains from taking action a_i net of the non-social cost of engaging in the action. s_i is i's private opinion about the desirability of a_i. $\tilde{s}_j$ is i's second-order belief, that is her belief about whether j privately endorses action a_i. $\widehat{\tilde{s}_-}$ is i's belief about the community's belief about the social norm $(\widehat{s_-})$, a proxy for the third-order beliefs we shift experimentally. i chooses $a_i = 1$ iff $E\left[U_w(1, s_i, \tilde{s}, \widehat{\tilde{s}_-})\right] > E\left[U_w(0, s_i, \tilde{s}, \widehat{\tilde{s}_-})\right]$. This means that we can characterize i's decision in terms of thresholds of $\gamma^*(s_i, \tilde{s}, \widehat{\tilde{s}_-})$ above which i choose action a_i for different sets of beliefs. We use this result to characterize the gap in the fraction of women who will act between women who believe $\tilde{s}_j = 1$ and those who believe $\tilde{s}_j = 0$ $(\Delta P(\widehat{\tilde{s}_-}))$. This behavior gap mimics the treatment effect in the active control design of our field experiment. This framework yields the following two main results:

> **Result 1: Beliefs about social norms affect behavior** A shift in second-order beliefs $\tilde{s}_j$ without shifting third-order beliefs $\widehat{\tilde{s}_-}$ leads to positive behavior gap $(\Delta P(\widehat{\tilde{s}_-}) > 0)$. That is, women who believe that the community is privately more supportive are more likely to act than those who believe that the community is privately opposed to their actions. This is due to the influence of $\tilde{s}_j$ on women's anticipated social consequences conditional on $a_i = 1$.

> **Result 2: Perceived mutual knowledge of social norms reinforces the effect of social norms** We define perceived mutual knowledge as the alignment of second and third-order beliefs (i.e., $\tilde{s}_j = \widehat{\tilde{s}_-}$). Perceived mutual knowledge of social norms increases the behavior in political activism between the cases of perceived private support and opposition to political activism $(\Delta P_{mut} > \Delta P(\widehat{\tilde{s}_-}))$. This happens because women

with $\tilde{s}_j = 1$ now always anticipate being praised and women with $\tilde{s}_j = 0$ now always anticipate punishment. Without perceived mutual knowledge and constant $\widehat{\tilde{s}_-}$ only one of these groups expects social consequences.

Our experiment is designed to directly test results 1 and 2 of the theoretical framework using an active control design combined with exogenous variation in mutual knowledge.

Strategic interactions We can extend our theoretical framework to allow for strategic interactions between political activists by setting $\gamma_i = \phi^k(\bar{a}) + \alpha_i$. $\phi^k(\bar{a})$ captures strategic interactions by making the return to political activism a function of the fraction of other women who choose $a_i = 1$ ($\bar{a}$). We analyze both the case of strategic complementarity $\phi^{comp}(\bar{a})' > 0$ and strategic substitutability $\phi^{sub}(\bar{a})' < 0$ as there is evidence for both patterns (e.g., Cantoni et al., 2019; González, 2020; Hager et al., 2022).

In Appendix Section C.3 we derive the following result about the role of strategic interactions

> **Result 3: Strategic complementarity increases the behavior gap while strategic substitutability decreases the behavior gap relative to the case of no strategic interactions.** Intuitively, strategic complementarity strengthens the effect of social norms because social norms affect beliefs about the actions of others in the same direction: an increase in the participation of others leads to an increased incentive to participate. Conversely, strategic substitutability counteracts the effects of social norms as an increase in the participation of others decreases the incentive to participate. Importantly, the behavior gap can even turn negative when strategic substitutability is strong enough.

We further show in the appendix that strategic complementarity increases the effect of mutual knowledge on the behavior gap. The effect of strategic substitutability on the sign

of the mutual knowledge effect is ambiguous. Mutual knowledge may increase or decrease the behavior gap depending on the strength of strategic motives.

3   Context and Design

Our study is set in rural Kyrgyzstan, a small open economy with lower-middle-income status. It has suffered several forced transitions of political power and multiple episodes of violent conflicts. Being the only (partial) democracy in the region, Kyrgyzstan has a tumultuous history of political violence and unrest. Since 1990, the country has seen two revolutions and two episodes of ethnic violence (Hager et al., 2019). Importantly, Kyrgyzstan is also an exemplary battleground for the international movement toward more female political empowerment. On the one hand, the country prides itself on its strong female role models during its nomadic days, but also during the USSR, where labor force participation was high. The country even had the only-ever female president in Central Asia in the early 2010s. On the other hand, a surge in Islamic radicalism coupled with deeply entrenched patriarchal values means women continue to be underrepresented in national and local constituencies.

While participation in national and local elections as voters is relatively balanced across genders, formal participation through other formal means is skewed towards men (Esenaliev and Kisunko, 2015). This is mirrored in women's economic participation with every second work-aged female being inactive in the labor market (Asian Development Bank, 2019). In practice, female political participation at local levels occurs through informal and formal channels and groups, such as women's committees, saving groups, and village health committees. We focus on such informal participation as our main outcome of interest as a natural behavioral outcome.

Norms around female political participation are shaped by two forces. On the one hand, there has to be a top-down push for more female representation. Kyrgyzstan introduced a 30% quota in the national parliament in 2007, and the same quota was introduced in

local councils in 2019. This increased the number of female members in local councils and generally made women more visible, potentially making female political participation more socially acceptable.

However, this push for female representation is opposed by an increasing spread of (conservative) Islam which promotes traditional gender norms (Heyat, 2004). This leads to an increasing pushback against female political (and also economic) participation in society. This is particularly relevant in rural areas like the ones our study is set in. Hence, there has been at least the perception of a shift away from social norms supportive of female political participation.

Jointly, these forces create an environment where social norms are changing but in an unclear direction. This uncertainty may give rise to misperceptions of prevailing social norms and even misperceptions of others' perceptions of social norms. Therefore, rural Kyrgyzstan is an excellent place to study whether (mis)perceptions of social norms and lack of mutual knowledge about norms influence female political activism.

3.1   Study design

We implement a field experiment providing information about social support for female political participation to 5,201 women across 150 villages in three provinces in Kyrgyzstan. Figure 2 displays the experimental design.

Sample   We recruit a sample of 5,201 women from rural subdistricts in three regions in Kyrgyzstan. The sample consists of a representative component of adult women in rural Kyrgyzstan and women who are particularly likely to be or consider being active in local politics. For the experimental part, we randomly sample around 30 female heads in each village through a random walk technique.[5]   In total, the representative component of the

---

[5]We randomly picked starting points and directions in each village and instructed enumerators to knock on each door until they reached their target number or ran out of time. If the door was not opened they

Figure 2: Experimental Design
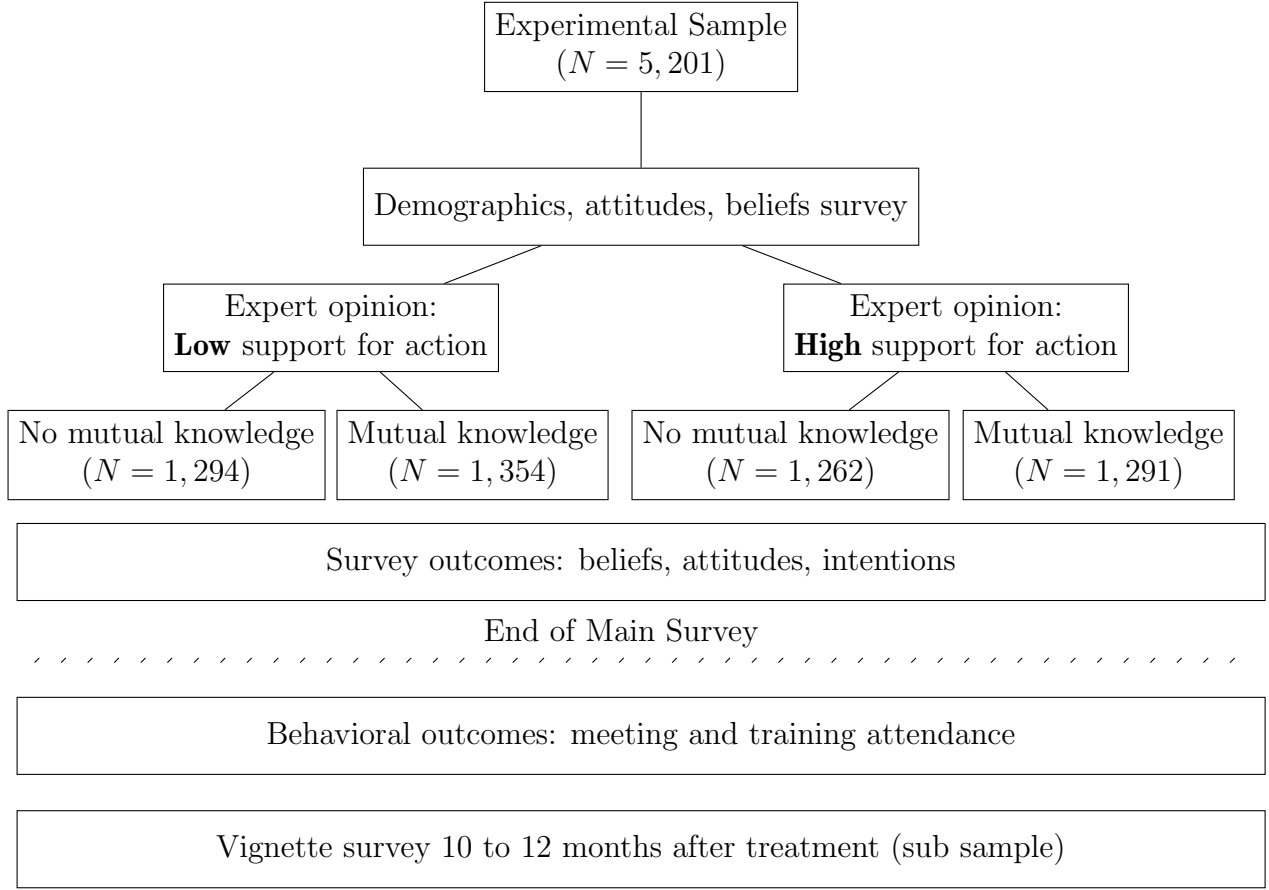
Notes: **Figure 2 illustrates the stylized experimental design of the field experiment.**

sample consists of 4,432.

The second group of women consists of 769 female members of the local Village Health Committee (VHCs).[6] We include this group of more socially and politically active women as we anticipated them to be more likely to attend the community meeting which constitutes

were instructed to continue onwards. If the door was opened, they were instructed to talk to the female head of the household. If the person was temporarily unavailable, they were instructed to return at a later point to conduct the interview.

[6]Village Health Committees are community groups made up of volunteers, mostly people who live in rural and semi-urban areas (Isabekova, 2021; Schueth, 2014). The Village Health Committees were established as part of the Swiss-funded Health Reform Support Project, implemented by the Swiss Red Cross from 2001 for over 17 years. The project aimed to enable rural communities to improve their health and to support the state healthcare system in partnering with these communities. Over 1600 VHCs were established at the peak with most of them being active after the completion of the Swiss project. Depending on the village's size, these groups usually have 5–10 members, comprising mostly women. The members are volunteers who raise awareness and promote health during the project and afterward. VHCs identify health problems in their villages and work on solutions.

our main behavioral outcome. VHC members sampled in the representative sample are attributed to this group. We pool both groups for the analysis to maximize power.

Once recruited, we conduct a baseline survey capturing demographic variables, general attitudes toward local and national politics, and beliefs about societal support for female political activism. This survey was either conducted in Kyrgyz or Russian, depending on the respondents' preferences. Table 1 displays the characteristics of our sample.[7] The women are, on average, 44 years old, have 3.75 children, and 86% are married. 96% belong to the ethnic Kyrgyz majority. Only 36% work to earn money with an additional 41% being housewives and 18% having retired. The women are relatively educated with 72% having completed secondary school and 25% having received tertiary education, which attests to Kirghiz Soviet Socialist Republic's focus on education.

Pre-Treatment Beliefs We also elicit women's higher-order beliefs about prevailing injunctive norms regarding female political participation. We elicit second-order beliefs, that is, beliefs about the private opinions of others, by asking "*Please think of 10 typical Kyrgyz [women / men]. What do you think? How many of these 10 [women / men] think that Kyrgyz women should participate more in local politics.*" Table 1 shows that, averaged across genders, we observe that women expect 56% of Kyrgyz people to want women to be more politically active. This is substantially higher for women (69%) compared to men (43%).

Second, we elicit women's third-order beliefs about injunctive norms. That is, we ask women to guess their neighbors' second-order beliefs. We ask the following after each of the second-order beliefs questions: "Now, imagine we ask the same question we just asked you to your neighbors. How do you think they would answer the question?"[8]

---

[7]Table A2 show how the representative and active sample of women differ in observable characteristics.

[8]This was more difficult to answer, as indicated by the increased number of missing values. To facilitate the understanding of the questions, the enumerators used visual aids presented in the Appendix (Figure A.1) and provided the opportunity to ask clarifying questions.

Table 1: Summary Statistics

	(1) Mean	(2) SD	(3) Min	(4) Max	(5) N
Panel A: Demographics					
Age	44.19	12.99	18	75	5201
Female	1.00	0.00	1	1	5201
Number of children	3.75	1.65	0	12	5201
Married	0.86	0.35	0	1	5201
Kyrgyz	0.96	0.21	0	1	5201
Village health committee member	0.15	0.36	0	1	5201
Housewife	0.41	0.49	0	1	5201
Working for money	0.35	0.48	0	1	5201
Retired	0.18	0.39	0	1	5201
Primary education	0.02	0.15	0	1	5127
Secondary education	0.72	0.45	0	1	5127
Tertiary education	0.25	0.43	0	1	5127
Panel B: Beliefs about support for female political activism					
Belief: % ave. support	55.99	17.83	0	100	5136
Neighbors' belief: % ave. support	50.71	20.74	0	100	4812
Belief: % women support	68.98	24.57	0	100	5171
Neighbors' belief: % women support	60.93	27.63	0	100	4897
Belief: % men support	43.03	21.32	0	100	5143
Neighbors' belief: % men support	40.62	22.43	0	100	4858
Panel C: Intentions and activism					
Likelihood of pol. action in next six months (0-4 Likert scale)	2.94	0.92	0	4	5201
Interest in community grant program (0-4 Likert scale)	2.86	0.90	0	4	5201

Notes: **Table 1 displays the characteristics of the women participating in the study.** The number of observations varies because of refusals (belief variables) or technical problems in early survey versions (education variables).

Table 1 shows that Kyrgyz women believe that neighbors are slightly more pessimistic than themselves, stating that neighbors believe that 51% women should be more active in politics. The difference is highly statistically significant ($p < 0.001$). The gap suggests that there is at least some degree of misalignment between second- and third-order beliefs, which we call a partial lack of perceived *mutual knowledge.* This lack of mutual knowledge is also present and significant when we consider gender-specific beliefs.

Figure 3: Opinions and higher order beliefs
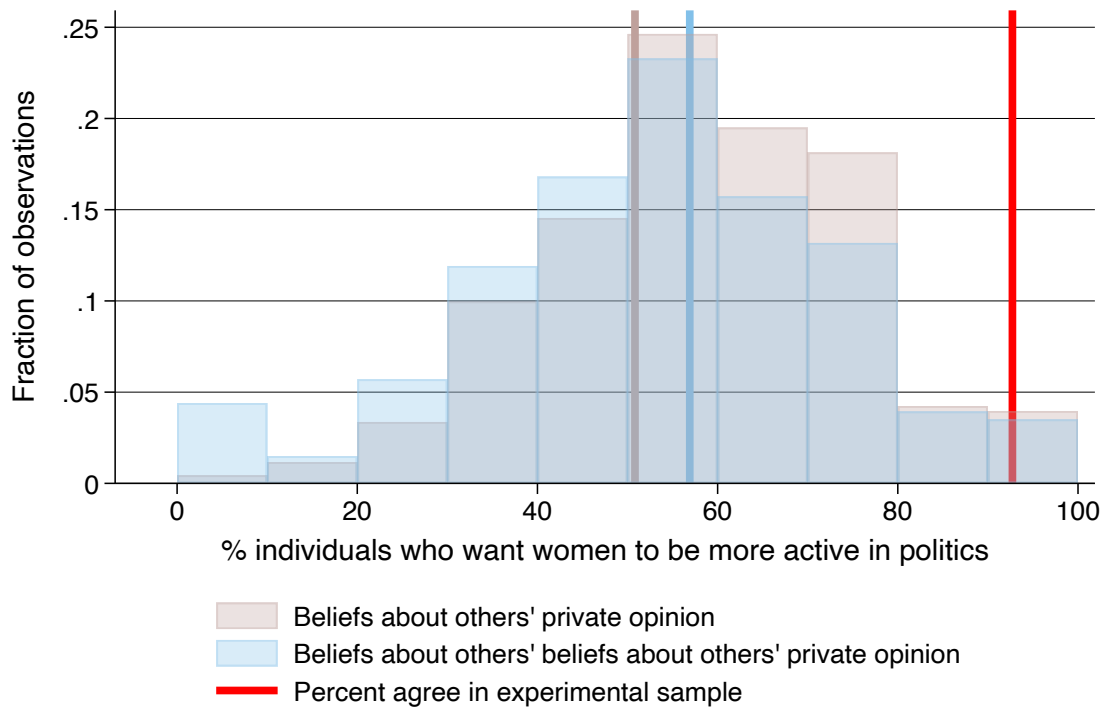
Notes: Figure 3 displays women's private support for political activism (red) line and their beliefs about others' private opinions as well as beliefs others' beliefs about support for female political activism.

Figure 3 plots women's beliefs about what others believe how many women out of 100 should be politically active as well as what others believe about other's opinions regarding female activism. There are two striking facts. First, the representative sample of women shows a strong conviction that many women should be active in politics: 92 percent (red bar). Second, the same women, however, believe that others are much less favorable toward female engagement. Specifically, women believe others only want 70 percent of women to be active. Third, even more striking, the sample of women thinks that others hold even more negative views of what others think regarding female engagement. On average, women believe that others think that others only want 62 percent of women to be active. The figure thus shows that as the order of beliefs increases, women become more skeptical. Put differently, there is *pluralistic ignorance* in that women think others think more negatively about female political engagement.

3.2 Experimental Variation in Information about Social Norms

Next, we provide participants with truthful information about social norms around female political participation. We cross-randomize two components to test whether (the lack of) mutual knowledge about social norms affects behavior. We randomize both treatments at the individual level using the random number generator included in SurveyCTO.[9]

Treatment 1: Providing Information about Social Norms First, we randomly provide women with either a high or a low *true* expert estimate about the fraction of Kyrgyz society who want women to be more active in politics. This design is similar to that used in Bursztyn et al. (2020a) but differs in two key aspects. First, we use an active control design to generate variation in second-order beliefs independent of participants' prior beliefs (Haaland et al.,

---

[9]The individual level randomization allows for the possibility of information spillovers across individuals which has the potential to attenuate treatment effects. However, our main behavioral outcome is measured only a few days after treatment, limiting the scope for such spillovers. Still, we acknowledge that our treatment effect estimates might underestimate the effect of treating everyone with the same information.

2023). Second, we provide information about injunctive rather than descriptive social norms. We provide information about injunctive norms as they matched the political discourse at the time. Kyrgyzstan has a 30 percent quota for female local councilors, and there was an active push by the government to get women involved in (local) politics. Injunctive norms reflect the degree to which this political agenda has been internalized by the Kyrgyz population.

We further randomize the gender source of the social norm information to test whether the source of social norms influences conformity. The enumerators read the following script and accompanied it by visual aids displayed in Figure A.2:

> "The research team consulted Kyrgyz experts on female political participation in Kyrgyzstan to understand whether people in Kyrgyzstan want women to be more active in politics. These experts have a lot of experience working or conducting research on female political participation. We like to tell you about what we found in our conversations with the experts. One expert told us that they think that [**High: 7** / **Low**: 1] out of 10 [women / men] believe that women should participate more in politics. In other words, a [**High: large majority** / **Low: small minority**] of [women / men] in Kyrgyzstan believe that women should become more active in politics."

The expert estimates are derived from a survey of 13 social scientists who work on Kyrgyzstan (most focus on gender and are Kyrgyz themselves). We rely on expert estimates to i) gain variation, and ii) because there were no reliable national-level estimates we could have provided to women. The low and high estimates we chose represent the lowest and highest estimates provided by the expert sample. The fact that we gave 1 (low) and 7 (high) out of 10, respectively, means that the low estimate, on average, provides more information relative to the average prior belief in the sample is 5.6. However, the active control design means that we estimate treatment effects for the second-order belief treatment by comparing the low and high groups rather than relying on the treatment

providing new information relative to prior beliefs.

Treatment 2: Creating Mutual Knowledge Next, we cross-randomize whether women are told that the same information will also be given to other community members. The enumerators read the following script (in Kyrgyz or Russian) and used the visual aids displayed in Figure A.3:

> We have conducted similar interviews across this and other communities. In the interviews, we told many of your neighbors that a [**large majority** / **small minority**] of [women / men] in Kyrgyzstan supports increased participation of women in politics. This means that many of your neighbors now know that there [**is** / **is no**] widespread support that women should become more active in politics.

This treatment is designed to create mutual knowledge about social norms by aligning second and third-order beliefs about social norms. It thus provides a direct test of Result 2 in the theoretical framework.

Vignette survey We conduct another survey with a subset of the sample 10 to 12 months after treatment as part of the companion study (Esenaliev et al., 2024). We attempt to recontact all women who attended the community grant meeting and a random subsample of women who did not attend the meeting. In total, we survey 3,760 women of whom 793 attended the community meeting. In this survey, we administer a series of vignettes that explore the mechanisms behind our main effect. We describe the vignettes and empirical findings in Section 6.

18

3.3 Outcome measurement

After the treatment administration, we immediately measure respondents' post-treatment beliefs about social norms in their community.[10] We ask two questions accompanied by the same visual aids as the pre-treatment beliefs. First, we ask: "Think of 10 typical [women / men] in your village. What do you think? How many of these 10 women believe that women in Kyrgyzstan should participate more in local politics?" Second, we ask the same question about how their neighbors would answer the previous question as we do pre-treatment. We also measure further post-treatment outcomes, including intended political participation.

Our main behavioral outcome is a dummy indicating whether a woman attended a meeting providing information about a community grant program held a few days after the baseline survey. This meeting was organized by a local NGO (Development Policy Institute) and the local village health committee. All women received information about the meeting time, location, and purpose at the end of the baseline survey. We recorded attendance at the meeting using sign-in sheets provided by the NGO.[11] On average, 17.5% of women in our sample attended this meeting.

Attendance at the community meeting is subject to social norms. During the vignette survey, we ask women how much they agree with the statement "If I attend such meetings, other villagers will know that I attended." Only 9% of respondents disagree or strongly disagree with this statement, while 67% agree or strongly agree. Moreover, women report explicit societal pressure in both directions: 28% state that they would be expected to not attend, and 57% state they would be expected to attend such community meetings. The observable nature of meeting attendance and the reported social pressure suggest that social norms and sanctions have the potential to shape attendance decisions in our context.

---

[10]We measure post-treatment beliefs about the community rather than Kyrgyzstan as a whole because we think that beliefs about local social norms are more decision-relevant and also likely to respond to the provided information (similar to Hager et al., 2023).

[11]This meeting also served to define the sample for a separate village-level intervention that provided group-efficacy training to village health committee members (Esenaliev et al., 2024).

4   The Effect of Information about Social Norms

We use the following pre-specified specification to estimate the effects of receiving a high expert forecast.[12]

$$y_i = \beta_0 + \beta_1 \text{High condition}_i + \delta_v + \Gamma X_i + \varepsilon_i \tag{2}$$

where y_i is the outcome of interest for individual i, High condition$_i$ indicates receiving the high expert opinion, δ_v represents village fixed effects. X_i represents pre-specified control variables that include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months. We estimate equation 2 for the sample without the mutual knowledge intervention. β_1 identifies the effect of receiving a high expert opinion rather than a low expert opinion without mutual knowledge.

Belief updating   Our intervention successfully leads women to update their beliefs about prevailing local norms. Women who receive a high expert opinion about prevailing support for female political activism have a 0.41 higher difference between post- and pre-treatment beliefs ($p < 0.001$; column 1 of Table 2). This is equivalent to a 4.1 percentage point increase in the perceived support for female political participation. Two-thirds of this effect is due to belief shifts in the low-support condition compared to one-third in the high-support condition, in line with the observation that average pre-treatment beliefs fall above the midpoint between the high and low-treatment information.

These results are robust to using belief levels and to disaggregating belief measures by gender (Table A3). The gender source of the treatment does not differentially affect average

---

[12]Our pre-analysis can be accessed at `https://aspredicted.org/dqm5-drwg.pdf`. Appendix Section D describes how we deviate from the pre-analysis plan

Table 2: Treatment Effects of Providing Information about High Support

	Δ Second-order belief	Attended meeting
	(1)	(2)
High condition	0.415***	-0.013
	(0.073)	(0.014)
Mean low condition	-0.292	0.193
Observations	2502	2556

Notes: **Table 2 shows that, while beliefs update, behavior is insensitive to providing information about others' private opinions.** *High condition* indicates the effect of receiving a high rather than a low expert opinion about the prevailing support for female political activism. Sample excludes the mutual knowledge treatment group. Column 1 shows the effect on the difference between post- and pre-treatment second-order beliefs (averaged across gender). Column 2 shows the effect on a dummy indicating attendance at a community meeting. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

belief updating (column 1 of Table A13). It does also not substantially affect the gap the treatment induces in gender-disaggregated beliefs (columns 2 and 3). This is evidence of substantial cross-updating across the gender domain of the intervention.

Behavior The belief updating does, on average, not translate into behavioral change. Column 2 of Table 2 shows that women in the high condition are not more likely to attend the community grant meeting. The effect is negative but relatively close to zero and not statistically significant (-1.3pp; $p = 0.32$).

The gender dimension of the treatment induces an insignificant but sizable difference (-3.3pp, $p = 0.23$); column 4 of Table A13). Accordingly, the social norm effects are more negative and almost significant for the female treatment condition (-3pp, $p = 0.12$) compared to the male treatment condition (0.3pp, $p = 0.87$).

Our theoretical framework suggests two factors that can explain the zero or even weakly

negative effects of the social norm treatment. First, a lack of perceived mutual knowledge about the provided information can attenuate the effect of the social norm treatment. Women might think that their community perceives social norms that do not align with their own beliefs. Hence, they might expect others in their community to react to their political activism based on their perceived social norms rather than their private opinion.[13] This channel can explain a small, insignificant treatment effect.

The second channel, strategic interactions between women, can explain the potential negative effects of the social norm treatment. This is because the treatment may lead to changes in anticipated attendance, as information about the private opinion of others is provided. If women exhibit strategic substitutability, similar to political activists in Cantoni et al. (2019) and Hager et al. (2023), this would further reduce the observed behavior gap. The notion is supported by the fact that the effect is somewhat more negative when the treatment is about women supports this notion. The meetings were targeted exclusively at women, hence, strategic interactions should only occur when women are perceived as more or less likely to attend.

However, if these strategic interactions are not too strong, mutual knowledge can still increase the behavior gap by overcoming the first of the two channels leading to a reduced social norm treatment effect.

5 The Effect of Perceived Mutual Knowledge

Does perceived mutual knowledge have a causal impact on the effectiveness of providing information about social norms? To tackle this question, we test whether the effect of the second-order beliefs treatment increases when the perception of mutual knowledge is created.

---

[13]Other effective interventions in the literature may inadvertently have created mutual knowledge through their experimental design. For example, respondents in Bursztyn et al. (2020a) take the survey in the same room at the same time. This might well lead participants to believe that others also receive the same information. However, they do not explicitly study mutual knowledge.

We estimate this Δ using the following regression equation:

$$y_i = \beta_0 + \beta_1 \text{High condition}_i + \beta_2 \text{Mutual knowledge}_i$$
$$+ \Delta \text{High condition}_i \times \text{Mutual knowledge}_i + \delta_v + \Gamma X_i + \varepsilon_i \tag{3}$$

where Mutual knowledge$_i$ is a dummy indicating the mutual knowledge treatment. We estimate the effect of the mutual knowledge treatment within each social norm condition using split sample regressions with the same controls. The full regression results are displayed in Table A5.

Belief updating We observe that women update their beliefs about their neighbors' beliefs about social support for female political activism more when they are told that their neighbors receive the same expert opinion (Table 3).[14] That is, creating perceived mutual knowledge leads to more updating of third-order beliefs. The mutual knowledge treatment leads to a 0.25 increase in the effect of the social norm treatment on third-order belief updating ($p = 0.026$; column 1). This is a 69% increase in the updating of third-order beliefs. The mutual knowledge treatment thus successfully causes belief updating in the intended direction.

The impact of perceived common knowledge on third-order beliefs is larger in the low-support condition compared to the high-support condition. Women in the low support condition who receive the mutual knowledge intervention decrease their difference in third-order beliefs by 0.16 more than women without this intervention ($p = 0.047$; column 2). However, women in the high support group see only a smaller and insignificant increase in updating in the mutual knowledge condition. They update their beliefs by 0.09 more, but this difference is not significant ($p = 0.276$; column 3). This is potentially due to cross-updating: Women in the high support group without mutual knowledge increase their post-treatment

---

[14]Table A4 shows that the belief updating is, to a larger extent, driven by updating about neighbors' beliefs about men's support for female political activism.

beliefs about neighbors' beliefs by 0.33, despite them not being told that other neighbors would receive the same information.

Table 3: The Effect of Perceived Mutual Knowledge

	Δ Third-order belief			Attended meeting		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.246** (0.111)			0.030* (0.018)		
Mutual knowledge		-0.164** (0.082)	0.080 (0.079)		-0.036*** (0.013)	-0.014 (0.013)
Mean no mutual knowledge	0.148	-0.029	0.330	0.185	0.193	0.176
Observations	4740	2410	2329	5201	2649	2552

Notes: **Table 3 shows that creating perceived mutual knowledge shifts behavior in the low, but not the high support condition.** Δ *High condition* indicates the difference of the second-order belief treatment effect between the mutual knowledge and no mutual knowledge condition. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment in sample split regressions. Columns 1 to 3 show the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism (averaged across gender). Columns 4 to 6 show the effect on a dummy indicating attendance at a community meeting. Columns 1 and 4 show how the effectiveness of the second-order belief treatment varies with the creation of perceived mutual knowledge. Columns 2 and 5 show the effect of perceived mutual knowledge in the low-support condition. Columns 3 and 6 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Behavior These belief effects of the mutual knowledge treatment translate into changes in behavior. The treatment effect on meeting participation of providing high relative to low expert opinions about social support for female political participation increases by 3pp when perceived mutual knowledge is created ($p = 0.098$; column 4). This is a sizable difference given the mean attendance in the non-mutual knowledge group of 18.8%. This result implies that providing information about social norms is more effective if recipients know or believe that others in their community receive the same information. This result provides causal

evidence supporting the theoretical argument that a lack of mutual knowledge can reduce the effectiveness of providing information about social norms.

In line with the effects on beliefs, we see that perceived mutual knowledge affects behavior most in the low support condition. Women in the low support condition reduce their meeting attendance by 3.6pp if they receive the mutual knowledge treatment ($p = 0.006$; column 5). This is a reduction in attendance of 17% relative to the no mutual knowledge mean. We observe no treatment effect of creating mutual knowledge in the high support condition. The treatment effect estimate is small (-0.014) and not statistically significant ($p = 0.31$; column 6). This asymmetry suggests that the effect of creating mutual knowledge is heterogeneous.

Heterogeneity by pre-treatment belief Can pre-treatment beliefs in our sample explain the difference in the mutual knowledge effect between the high and low conditions? We observe that women's pre-treatment second- and third-order beliefs are, on average, closer to the high support information. This plausibly leads to the reduced belief updating we observe in the high condition. To understand whether the reduced belief updating drives the lack of behavioral change induced by mutual knowledge, we analyze treatment effect heterogeneity by how much information we provide relative to women's pre-treatment third-order beliefs. We define their third-order *updating potential* as the difference between individuals' pre-treatment third-order beliefs and the provided expert opinion. Women with positive updating potential would interpret the mutual knowledge condition as a positive shock to their neighbors' beliefs. Conversely, women with negative updating potential would interpret the mutual knowledge condition as a negative shock to their neighbors' beliefs.

Figure 4 shows that the effect of the mutual knowledge treatment is almost entirely driven by women with negative updating potential. That is, women who are told that their neighbors receive information that is below their presumed beliefs are then less likely to attend the meeting compared to women not receiving the mutual knowledge treatment.

On the contrary, women who anticipate a positive shock to their neighbors' beliefs do not change their behavior. This broad pattern is also reflected when we consider each treatment condition separately, though the range of updating potential is more constrained (Figure A.4). This suggests that the asymmetry in the effect of the mutual knowledge treatment does not exclusively arise because of the distribution of pre-treatment beliefs.[15] Rather, the results suggest that there is a fundamental difference between creating mutual knowledge about a positive or negative norm.[16]

Gender source of norms Social reactions to female political activism do not depend on the average social norms but on prevailing social norms in relevant subgroups. For example, political activists who have closer ties with other activists are less likely to exhibit free-riding behavior in response to being told about high descriptive norms (Hager et al., 2023). One particularly salient dimension is whether we provide information about social norms among men or women. It is ex-ante unclear which group should elicit a stronger response. On the one hand, men are more likely to be in positions of power so their beliefs and actions might influence women's decisions more (e.g., Cheema et al., 2023). On the other hand, women tend to form closer social connections among themselves making social punishment or praise more important to our respondents.

The effect of mutual knowledge does not differ significantly between providing information about social norms among women and men (Table A7). However, point estimates of the effect of mutual knowledge are larger and more significant when the provided information concerns women rather than men. Providing information about women rather than men leads to 219% larger differences in belief updating, though differences are not statistically

---

[15]It is possible for the mutual knowledge treatment to affect behavior through other channels than changes in belief levels. For example, the treatment might decrease the uncertainty associated with the beliefs about neighbors' beliefs. Unfortunately, we do not measure belief uncertainty.

[16]Results for intentions are qualitatively consistent but less precisely estimated (Table A6). The intention measures likely suffering from social desirability bias. For example, 88% of women in the no mutual knowledge group that that they plan to attend the meeting, whereas only 18.5% end up going. This may explain why the intention effects are less precisely estimated. The relative unreliability of intention measures is also found in other studies in the literature (e.g., Hager et al., 2021).

Figure 4: Treatment Effects by Updating Potential
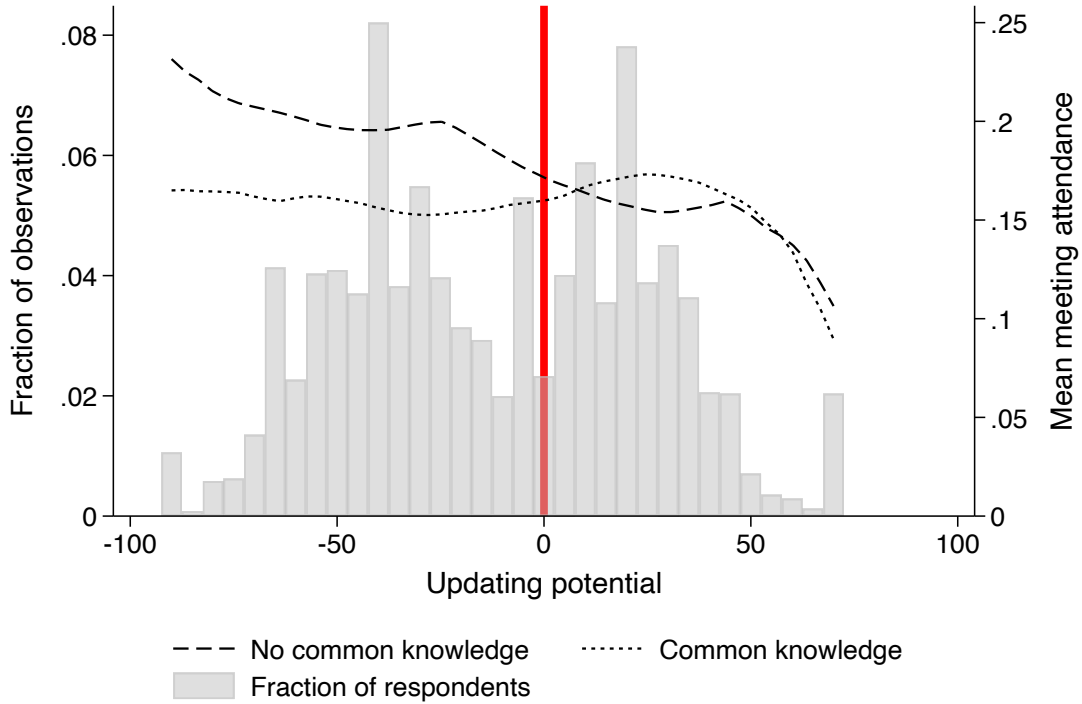
Notes: **Figure 4 shows that the mutual knowledge treatment only affects meeting attendance for women with negative updating potential.** Updating potential displayed on the x-axis is defined as pre-treatment third-order belief minus the provided information about societal support for female political activism. Women with positive updating potential would interpret the mutual knowledge condition as a positive shock to their neighbors' beliefs. Conversely, women with negative updating potential would interpret the mutual knowledge condition as a negative shock to their neighbors' beliefs. Black lines represent local polynomial regressions of meeting attendance on updating potential using an Epanechnikov kernel with bandwidth 15.

significant.[17] Differences are similarly sizable but insignificant for both the low and high conditions (columns 2 and 3).

Mutual knowledge makes the social norm treatment 4.9pp effective when information about women is provided ($p = 0.059$) whereas it only increases by 1.2pp when information about men is provided ($p = 0.651$). However, the difference between both is not quite significant at $p = 0.316$ (column 4). The difference is induced by a relatively large difference in the effect of mutual knowledge in the high condition (2.1pp, $p = 0.435$) rather than differences in the low condition (0.8pp, $p = 0.764$).

Overall, we find at most suggestive evidence that providing information about women rather than men may enhance the effect of mutual knowledge on behavior. However, the heterogeneity is relatively imprecisely estimated and we caution against drawing strong conclusions from our analysis.

Medium-term effects   While we observe two longer-term behavioral outcomes[18], our experiment is not designed to study longer-term treatment effects. Instead, it is designed to provide well-powered evidence that higher-order beliefs can affect (political) behavior. We randomize our information treatment at the individual level. This means that there are likely to be substantial information spillover effects over time. Given that we provide contradicting expert opinions to different women, it is plausible that, over time, women's information sets converge to an intermediate point. Such spillovers can occur through direct communication or through observing the behavior of others. Our measurement further creates an opportunity for such inference by providing a relatively observable

---

[17]Table A12 shows that updating of gendered beliefs is mostly congruent with treatment but generally stronger for information and beliefs about women.

[18]We observe whether women attended the four collective efficacy training sessions for a random half of the villages where the intervention of Esenaliev et al. (2024) was implemented. These sessions take place about one month apart and start about one month after the community grants meeting. Second, we also observe whether women submit a grant application to the community grant program. The program starts accepting applications after all training sessions are completed, roughly six months after the initial meeting. The application window is open for two months.

opportunity for women to engage in local politics. We thus do not expect treatment effects to last.

Indeed, we observe that the effect of the mutual knowledge treatment on the effectiveness of providing social information does not persist in the medium term (columns 1 and 4 of Table 3). Effects in the low group initially persist but then become insignificant. The estimated effect of the mutual knowledge treatment in the high condition remains negative and even is significant for the longest-term outcome we observe.

6 Social Punishment and Praise as Mechanism

We use vignette experiments to provide evidence about the mechanisms through which perceived mutual knowledge affects behavior. We interview 3,760 women in the same communities 10 to 12 months after the first survey and present them with a series of scenarios and choices. The sample of women is broadly similar to the field experiment (column 1 of Table A9).

We first explore the role of social punishment and praise in driving behavior by asking women how they would react to others' meeting attendance conditional on perceived social norms and private support for political activism. That is, we ask them about the second stage of the decision process in the theoretical model.[19] We start by introducing the scenario: *Imagine you lived in a village and believed that women in Kyrgyzstan should [**not**] be more active in politics. However, you also believe that other villagers are [**not**] supportive of women being active in politics. You see that these women attended the meeting. If given the chance, would you want to punish them or praise them for attending the meeting? You can also decide to do nothing.*

---

[19]We focus on women's own praise and punishment behavior rather than on their beliefs about punishment and praise for their own actions because pilot results show that hypothetical vignettes with third-order beliefs were difficult to comprehend for respondents.

We mimic the field-experimental variation by cross-randomizing whether the acting women in the scenario are supportive themselves and whether social norms are perceived as supportive or not. The treatment groups are balanced on observables (Table A9). We analyze the data in parallel to the main experiment.[20]

Table 4 shows evidence that mutual knowledge about social norms affects how women react to others' meeting attendance.[21] Mutual knowledge decreases praise and increases punishment when social norms oppose female political activism (columns 2 and 5). However, the effect, relative to the no mutual knowledge mean, is substantially larger for punishment (266%, $p = 0.013$) than for praise (7.2%, $p = 0.093$).[22]

In line with this pattern, we observe that mutual knowledge only significantly affects the difference in punishment between the supportive and opposing scenarios: It reduces the difference in punishment behavior by 1.4pp ($p = 0.013$), which again is a large increase of 180% relative to the control mean. On the other hand, mutual knowledge does not significantly affect the difference in praise between the two scenarios ($p = 0.725$). While the point estimate is 1.1pp, the relative effect size is very small at only 2%. In balance, we interpret these results as suggestive evidence that mutual knowledge makes the social norm intervention more effective because the likelihood of receiving punishment increases when there is mutual knowledge about norms opposing female political activism.

Perception of attendants   These effects on social reactions are accompanied by a shift in the perceived virtue of women attending the meeting in the vignettes. Table 5 shows that mutual knowledge affects how perceptions of meeting attendants differ between low and high-support scenarios. Mutual knowledge makes the effect of the social norm treatment on

---

[20]We make slight changes in the control variables due to measurement differences between surveys. This does not affect the results beyond small changes in point estimates and standard errors.

[21]Table A10 shows that reactions to meeting non-attendance are mostly insensitive to mutual knowledge.

[22]In absolute terms, the effect on praise is larger at 3.8pp against 1.1pp for punishment. However, respondents are, on average, much more likely to state that they would praise attendants (53%) rather than punish them (less than 1%). This stark difference is likely due to women's personal convictions; only 6.73% of respondents do not want women to be more politically active.

Table 4: Vignette Experiment: Reactions to Meeting Attendance

	Praise attendance			Punish attendance		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.011 (0.032)			-0.014** (0.005)		
Mutual knowledge		-0.038* (0.023)	-0.021 (0.024)		0.011** (0.005)	-0.000 (0.003)
Mean no mutual knowledge	0.529	0.530	0.528	0.004	0.003	0.004
Observations	3760	1940	1820	3760	1940	1820

Notes: **Table 4 shows that perceived mutual knowledge affects social reactions to meeting attendance in vignette experiments.** The vignette experiments create variation that mimics the variation created in the main experiment. Columns 1 to 3 show effects on a dummy indicating that individuals would praise other women for attending a community meeting. Columns 4 to 6 show effects on a dummy indicating that individuals would punish other women for attending a community meeting. Columns 1 and 4 show how the difference between supportive and opposing norms around female political activism differs by mutual knowledge. Columns 2 and 5 show the effect of mutual knowledge when opposing norms are prevalent. Columns 3 and 6 show the effect of mutual knowledge when supportive norms are prevalent. Control variables include age, a self-employment dummy, a wage employment dummy, a dummy for being married, a dummy for being a village health committee member, dummies for having secondary, and tertiary education, the standardized support for female political activism, and measures of network size and structure (# friends, # friend talk about politics, # friends whose opinion matter), baseline treatment status, past meeting attendance, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

the fraction of women who see attendants as good community members 12.1pp ($p < 0.001$) larger. It also reduces the effect on the fraction of women who perceive meeting attendants as bad community members by 1.9pp ($p = 0.004$). While the latter effect is smaller in absolute terms, it is again substantially larger in relative terms (17% vs 280% of the now mutual knowledge mean). In line with the increase in punishment, the mutual knowledge effect on negative perceptions is driven by the low support condition.

Mutual knowledge also increases the fraction of women who perceive attendants as good community members by 6.8pp ($p = 0.001$). However, this does not translate into more praise (Table 4, column 3). This suggests that changes in social stigma or virtue are not sufficient to enforce norms, as commonly assumed in models of social norms (e.g., Bénabou and Tirole, 2006; Bursztyn et al., 2023).

Visibility of social reaction Social punishment is seen as more visible than praise. It is also mostly social and not economic. To assess the nature of social reactions, we ask women to select how they think that others would punish and praise them. Figure A.5, Panel A shows that women expect punishment to mostly consist of public verbal scorn (44%) and to a lesser extent private verbal scorn (20%). Fewer women select having a bad attitude (17%) or publicly being spoken badly about (14%) as a type of punishment. Only 6% select any of the economic reactions. The type of praise is also not economic in nature. This is similar for praise (Panel B). Only 8% select one of the economic options. However, responses are evenly distributed between private and public verbal praise (34% for both), suggesting that the decision to socially punish others might be more visible and hence more subject to social norms relative to social praise. This is in line with mutual knowledge about low support for female political activism being more effective in changing punishing rather than praising behavior.

Overall, the vignette evidence suggests the following causal chain affecting political activism. Mutual knowledge of negative social norms increases negative perceptions of

Table 5: Vignette Experiment: Perception of Meeting Attendants

	Good community members			Bad community members		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.121*** (0.029)			-0.019*** (0.007)		
Mutual knowledge		-0.060*** (0.021)	0.068*** (0.021)		0.020*** (0.006)	0.002 (0.004)
Mean no mutal knowledge	0.727	0.716	0.738	0.005	0.004	0.005
Observations	3760	1940	1820	3760	1940	1820

Notes: **Table 5 shows that perceived mutual knowledge affects how women attending the meeting are perceived.** The vignette experiments create variation that mimics the variation created in the main experiment. Columns 1 to 3 show effects on a dummy indicating that respondents perceive women attending the meeting as good members of the community. Columns 4 to 6 show effects on a dummy indicating that respondents perceive women attending the meeting as bad members of the community. Columns 1 and 4 show how the difference between supportive and opposing norms around female political activism differs by mutual knowledge. Columns 2 and 5 show the effect of mutual knowledge when opposing norms are prevalent. Columns 3 and 6 show the effect of mutual knowledge when supportive norms are prevalent. Control variables include age, a self-employment dummy, a wage employment dummy, a dummy for being married, a dummy for being a village health committee member, dummies for having secondary, and tertiary education, the standardized support for female political activism, and measures of network size and structure (# friends, # friend talk about politics, # friends whose opinion matter), baseline treatment status, past meeting attendance, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

politically active women. This in turn leads to an increase in social punishment for being politically active, which leads to reduced political activity. Mutual knowledge of positive social norms increases positive perceptions of politically active women. This does not lead to an increase in expected social praise for being politically active, potentially because the action of praise is more private than punishment and thus not subject to social norms. Ultimately, political activity is then also not affected.

Persuasion effects It is possible that mutual knowledge about social norms persuaded women to change their private opinions about female political activism. This could in turn affect activism and drive the observed treatment effects. To analyze this potential mechanism, we use the fact that 2897 women were interviewed in both the vignette survey and the sample of treated women. In the vignette survey, we also ask women's first-order beliefs, that is whether they agree that "women in Kyrgyzstan should be more active in politics" on a five-point Likert scale. We define agreement with the norm as either "strongly agreeing" or agreeing" with the statement, while "strongly disagree" or "disagree" is coded as agree. We then estimate whether the mutual knowledge treatment affects agreement with the social norm.

We find some evidence that the mutual knowledge treatment shifts private opinions. We find that mutual knowledge increases the social norm effect on agreement by an insignificant 3.6pp ($p = 0.273$, column 1 of Table A11). However, we find that mutual knowledge about low social norms decreases agreement with the social norm by 5.1pp ($p = 0.038$, column 2) while we observe no significant effect of mutual knowledge in the high condition (column 3). We similarly observe no effects on disagreement with the norm in either condition with very small point estimates (columns 4 to 6).

While this pattern could potentially explain the treatment effects on meeting attendance, we find only limited evidence that persuasion mediates the mutual knowledge effect. We first confirm that the treatment effects on meeting attendance are also present in

the subsample of women interviewed in both surveys: Columns 7 to 9 of Table A11 show similar treatment effects as in the full sample. To test whether changes in private opinions mediate the effects, we control for women's agreement with the social norm in our main treatment effect regressions (Columns 10 to 12). First, we find that agreement with the social norm is significantly associated with between 15 and 19pp higher meeting attendance. However, the estimated effect of mutual knowledge only reduces by between 12% and 15%. This suggests that persuasion only plays a minor role in explaining the effect of mutual knowledge.

7 Conclusion

In this paper, we show that mutual knowledge about prevailing social norms can be an important factor in shaping political behavior. Creating mutual knowledge increases the effect of a typical social norm information intervention significantly. We provide suggestive evidence that this is due to the fear of social punishment rather than the possibility that social praise drives political behavior.

What can we learn about the role of higher-order beliefs about social norms outside of the context of our study? First, the key premise of typical social norm interventions is that there are misperceptions about social norms, though academic studies with active control designs do not rely on this assumption. Such misperceptions are more likely to arise in contexts with shifting or competing social norms, where individuals have had limited time to adjust their perceptions. Our study emphasizes that misperceptions that go beyond 'pluralistic ignorance' about others' private opinions could slow down changes in social norms. This is supported by our evidence that, while there is some cross-updating, privately delivered social norm interventions do not shift higher-order beliefs to the same extent. Hence, there is scope for higher-order beliefs to affect behavior, as we document in this study.

Second, we find that higher-order beliefs matter differentially for positive and negative

norms. They reinforce the effect of negative norms while not affecting the effect of positive norms, likely because of differences in how norms are enforced through social sanctioning. These findings caution that the nature of norm enforcement is a crucial determinant of the effect of higher-order beliefs. However, there is evidence that negative sanctions are generally more commonly used to enforce social norms (Fehr and Schurtenberger, 2018), suggesting that our findings might be more widely applicable.

Finally, we have no evidence about the longer-term effects of shifting higher-order beliefs about social norms. Our individual-level randomization limits our ability to study longer-term effects and the dynamics of local social norms. Existing work on public signals about social norms suggests that sudden shifts in perceptions and behavior are possible, this work does not explicitly test for the role of higher-order beliefs (e.g., Bursztyn et al., 2020b). Future work could go further by randomizing at the village level and providing local information and by studying how local social norms and their perceptions shift over time. Similarly, engineering of common instead of just mutual knowledge about social norms, for example through community meetings as in Gottlieb (2016b), could further strengthen the effect of social norm treatments.

Our findings have practical implications for both researchers and policymakers. For policymakers, our results highlight the need to take higher-order beliefs seriously when trying to shift social norms through targeted information interventions. Our results suggest that creating mutual knowledge of negative norms can magnify the effect of typical social norm interventions. However, our findings also highlight that shifting higher-order beliefs may only work if they go along with shifts in the anticipated social reactions. Strategic substitutability of actions can reduce or potentially even reverse the potential effect of shifting higher-order beliefs.

For researchers attempting to shift social norms, our findings suggest that they should measure whether their interventions also shift higher-order beliefs as the lack of such shifts

may be crucial to explaining the (in)effectiveness of information experiments. Ideally, experiments should be designed to deliberately manipulate mutual or even common knowledge and include a discussion of this design aspect.

References

Abbink, Klaus, Lata Gangadharan, Toby Handfield, and John Thrasher, "Peer punishment promotes enforcement of bad social norms," *Nature communications*, 2017, *8* (1), 609.

Allcott, Hunt, "Social norms and energy conservation," *Journal of public Economics*, 2011, *95* (9-10), 1082–1095.

Andreoni, James, Nikos Nikiforakis, and Simon Siegenthaler, "Predicting social tipping and norm change in controlled experiments," *Proceedings of the National Academy of Sciences of the United States of America*, 2021, *118* (16).

Angeletos, George-Marios and La'O Jennifer, "Incomplete information, higher-order beliefs and price inertia," *Journal of Monetary Economics*, 2009, *56*, S19–S37.

Asian Development Bank, "2019 Annual Report," 2019.

Banerjee, Snehal and Ilan Kremer, "Disagreement and learning: Dynamic patterns of trade," *The Journal of Finance*, 2010, *65* (4), 1269–1302.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American economic review*, 2006, *96* (5), 1652–1678.

Bernhard, Helen, Ernst Fehr, and Urs Fischbacher, "Group affiliation and altruistic norm enforcement," *American Economic Review*, 2006, *96* (2), 217–221.

Bicchieri, Cristina, *The grammar of society: The nature and dynamics of social norms*, Cambridge University Press, 2005.

— , *Norms in the wild: How to diagnose, measure, and change social norms*, Oxford University Press, 2016.

Brandon, Alec, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer, "Testing for crowd out in social nudges: Evidence from a natural field

experiment in the market for electricity," *Proceedings of the National Academy of Sciences of the United States of America*, 2019, *116* (12), 5293–5298.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

_ , **Georgy Egorov, and Stefano Fiorin**, "From extreme to mainstream: The erosion of social norms," *American economic review*, 2020, *110* (11), 3522–3548.

_ , _ , **Ingar Haaland, Aakaash Rao, and Christopher Roth**, "Justifying Dissent," *Quarterly Journal of Economics*, 2023, *138* (3), 1403–1451.

_ , **Thomas Fujiwara, and Amanda Pallais**, "'Acting wife': Marriage market incentives and labor market investments," *American Economic Review*, 2017, *107* (11), 3288–3319.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong, "A Cognitive Hierarchy Model of Games," *The Quarterly Journal of Economics*, 08 2004, *119* (3), 861–898.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement," *Quarterly Journal of Economics*, 2019, *134* (2), 1021–1077.

Cespa, Giovanni and Xavier Vives, "The beauty contest and short-term trading," *The Journal of Finance*, 2015, *70* (5), 2099–2154.

Cheema, Ali, Sarah Khan, Asad Liaqat, and Shandana Khan Mohmand, "Canvassing the Gatekeepers: A Field Experiment to Increase Women Voters' Turnout in Pakistan," *American Political Science Review*, 2023, *117* (1), 1–21.

Chen, Yan, F Maxwell Harper, Joseph Konstan, and Sherry Xin Li, "Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens," *American Economic Review*, 2010, *100* (4), 1358–1398.

Coibion, Olivier, Yuriy Gorodnichenko, Saten Kumar, and Jane Ryngaert, "Do you know that I know that you know. . . ? Higher-order beliefs in survey data," *The Quarterly Journal of Economics*, 2021, *136* (3), 1387–1446.

Costa-Gomes, Miguel, Vincent P Crawford, and Bruno Broseta, "Cognition and behavior in normal-form games: An experimental study," *Econometrica*, 2001, *69* (5), 1193–1235.

Crawford, Vincent P. and Nagore Iriberri, "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica*, 2007, *75* (6), 1721–1770.

_ , **Miguel A. Costa-Gomes, and Nagore Iriberri**, "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, March 2013, *51* (1), 5–62.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta, "A survey of experimental research on contests, all-pay auctions and tournaments," *Experimental Economics*, 2015, *18*, 609–669.

Ensminger, J. and J. Henrich, *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective* Russell Sage Foundation Series on Trust, Russell Sage Foundation, 2014.

Esenaliev, Damir and Gregory Kisunko, "Local budget transparency and participation: evidence from the Kyrgyz Republic," *World Bank Policy Research Working Paper*, 2015, (7154).

_ , **Anselm Hager, Lukas Hensel, and Elnura Kazakbaeva**, "Collective Efficacy and Female Political Engagement: Evidence from a Field Experiment in Kyrgyzstan," *In progress*, 2024.

Fehr, Ernst and Ivo Schurtenberger, "Normative foundations of human cooperation," *Nature human behaviour*, 2018, *2* (7), 458–468.

＿ **and Urs Fischbacher**, "Third-party punishment and social norms," *Evolution and human behavior*, 2004, *25* (2), 63–87.

Frey, Bruno S. and Stephan Meier, "Social Comparisons and Pro-Social Behavior: Testing Conditional Cooperation in a Field Experiment," *American Economic Review*, 2004, *94* (5), 1717–1722.

George, Siddharth, Sarika Gupta, and Yusuf Neggers, "Can We Text Criminal Politicians Out of Office? Evidence from a Mobile Experiment in India," *Working Paper*, 2019.

Gerber, Alan S, Donald P Green, and Christopher W Larimer, "Social pressure and voter turnout: Evidence from a large-scale field experiment," *American political Science review*, 2008, *102* (1), 33–48.

Giné, Xavier and Ghazala Mansuri, "Together We Will: Experimental Evidence on Female Voting Behavior in Pakistan," *American Economic Journal: Applied Economics*, 2018, *10* (1), 207–235.

González, Felipe, "Collective action in networks: Evidence from the Chilean student movement," *Journal of Public Economics*, 2020, *188*, 104220.

Gorodnichenko, Yuriy and Xiao Yin, "Higher-Order Beliefs and Risky Asset Holdings," *NBER Working Paper 32680*, 2024.

Gottlieb, Jessica, "Common Knowledge and Voter Coordination: Experimental Evidence from Mali," in André Blais, Jean François Laslier, and Karine Van der Straeten, eds., *Voting Experiments*, Switzerland: Springer International Publishing, 2016, pp. 89–113.

＿ , "Why Might Information Exacerbate the Gender Gap in Civic Participation? Evidence from Mali," *World Development*, 2016, *86*, 95–110.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing Information Provision Experiments," *Journal of Economic Literature*, 2023, *61* (1), 3–40.

Hager, Anselm, Krzysztof Krakowski, and Max Schaub, "Ethnic Riots and Prosocial Behavior: Evidence from Kyrgyzstan," *American Political Science Review*, 2019, *113* (4), 1029–1044.

_ , **Lukas Hensel, Johannes Hermle, and Christopher Roth**, "Does Party Competition Affect Political Activism?," *Journal of Politics*, 2021, *83* (4), 1681–1694.

_ , _ , _ , **and** _ , "Group Size and Protest Mobilization across Movements and Countermovements," *American Political Science Review*, 2022, *116* (3), 1051–1066.

_ , _ , _ , **and** _ , "Political Activists as Free-Riders: Evidence from a Natural Field Experiment," *The Economic Journal*, 2023, *133* (653), 2068–2084.

Hallsworth, Michael, John A List, Robert D Metcalfe, and Ivo Vlaev, "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance," *Journal of public economics*, 2017, *148*, 14–31.

Heyat, Farideh, "Re-Islamisation in Kyrgyzstan: gender, new poverty and the moral dimension," *Central Asian Survey*, 2004, *23* (3-4), 275–287.

Huo, Zhen and Naoki Takayama, "Rational Expectations Models with Higher-Order Beliefs," *Review of Economic Studies*, 2024, p. rdae096.

Isabekova, Gulnaz, "Mutual learning on the local level: The swiss red cross and the village health committees in the Kyrgyz Republic," *Global Social Policy*, 2021, *21* (1), 117–137.

Jayachandran, Seema, "Social Norms as a Barrier to Women's Employment in Developing Countries," *IMF Economic Review*, 2021, *69* (3), 576–595.

Milazzo, Annamaria and Markus Goldstein, "Governance and women's economic and political participation: Power inequalities, formal constraints and norms," *World Bank Research Observer*, 2019, *34* (1), 34–64.

Nagel, Rosemarie, "Unraveling in guessing games: An experimental study," *The American economic review*, 1995, *85* (5), 1313–1326.

Nimark, Kristoffer, "Dynamic pricing and imperfect common knowledge," *Journal of monetary Economics*, 2008, *55* (2), 365–382.

Paxton, Pamela, Sheri Kunovich, and Melanie M. Hughes, "Gender in politics," *Annual Review of Sociology*, 2007, *33* (June 2015), 263–284.

Perez-truglia, Ricardo and Guillermo Cruces, "Partisan Interactions," *Journal of Political Economy*, 2017, *125* (4), 1208–1243.

Prillaman, Soledad Artiz, "Strength in Numbers: How Women's Groups Close India's Political Gender Gap," *American Journal of Political Science*, 2021, *0* (0), 1–21.

Schueth, Tobias, "The Community Action for Health Programme in Kyrgyzstan: A countrywide partnership for health promotion between the health system and communities.," *International Journal of Integrated Care (IJIC)*, 2014, *14.*

Shang, Jen and Rachel Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The Economic Journal*, 2009, *119* (540), 1422–1439.

A   Appendix Figures

Figure A.1: Visual Aids for Belief Elicitation
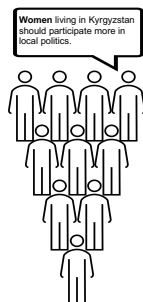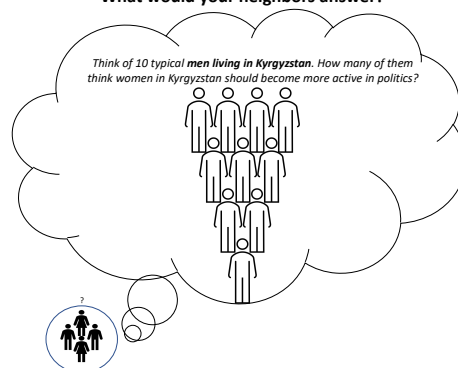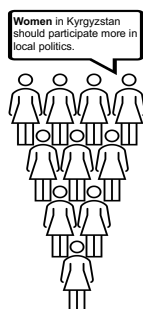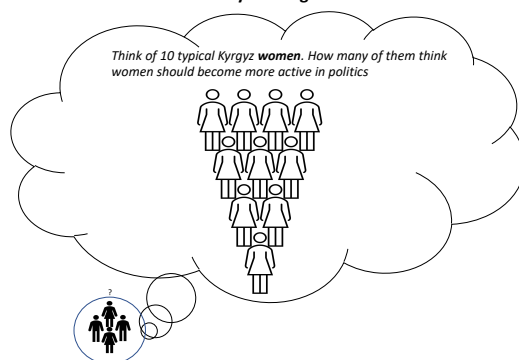
Notes: **Figure A.1 shows the visual aids used to elicit pre- and post-treatment beliefs about support for female political participation.** The visual aids were shown when enumerators read the text described in Section 3.1.

Figure A.2: Visual Aids for Second-Order Belief Treatment

Men, low

Men, high

Women, low

Women, high
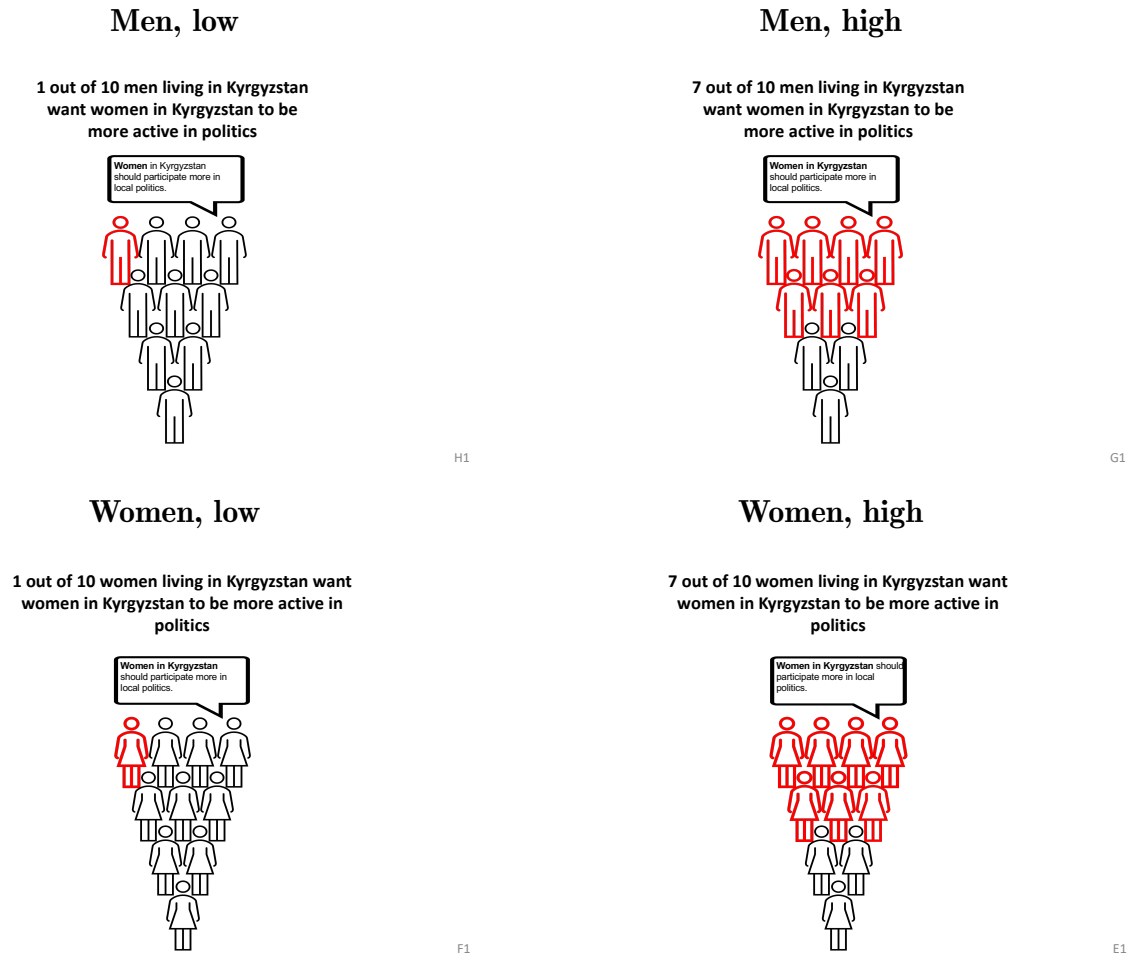
Notes: **Figure A.2 the visual aids used to help communicate the second-order belief information treatment.** The visual aids were shown when enumerators read the text described in Section 3.1.

Figure A.3: Visual Aids for Mutual Knowledge Belief Treatment

Men, low

We told many of your neighbors that a small minority of MEN in Kyrgyzstan support increased participation of women in politics.

Men, high

We told many of your neighbors that a large majority of MEN in Kyrgyzstan support increased participation of women in politics.

Women, low

We told many of your neighbors that a small minority of WOMEN in Kyrgyzstan support increased participation of women in politics.

Women, high

We told many of your neighbors that a large majority of WOMEN in Kyrgyzstan support increased participation of women in politics.

Notes: **Figure A.2 the visual aids used to help communicate the information treatment.** The visual aids were shown when enumerators read the text described in Section 3.1.

Figure A.4: Treatment Effects by Updating Potential and Condition

Panel A: Low support condition

Panel B: High support condition
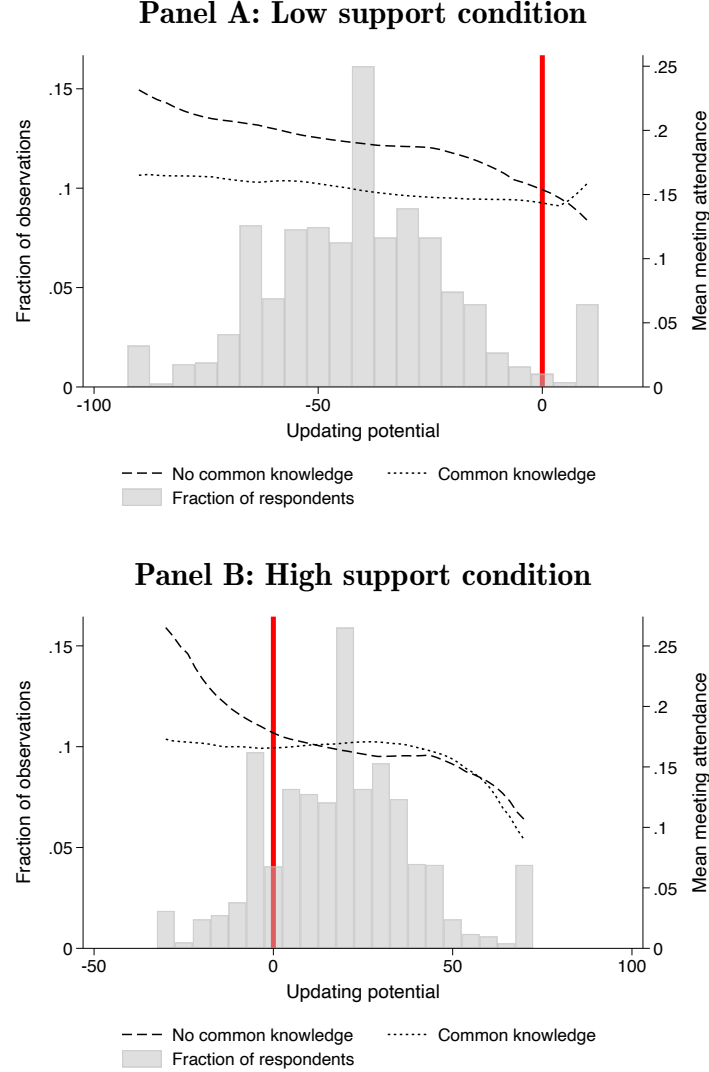
Notes: **Figure A.4 shows that the mutual knowledge treatment only affects meeting attendance for women with negative updating potential.** Panel A uses only women in the low support condition. Panel B uses only women the high support condition. Updating potential displayed on the x-axis is defined as pre-treatment third-order belief minus the provided information about societal support for female political activism. Women with positive updating potential would interpret the mutual knowledge condition as a positive shock to their neighbors' beliefs. Conversely, women with negative updating potential would interpret the mutual knowledge condition as a negative shock to their neighbors' beliefs. Black lines represent local polynomial regressions of meeting attendance on updating potential using an Epanechnikov kernel with bandwidth 1.5).

Figure A.5: Expected Type of Social Reactions

Panel A: Type of punishment
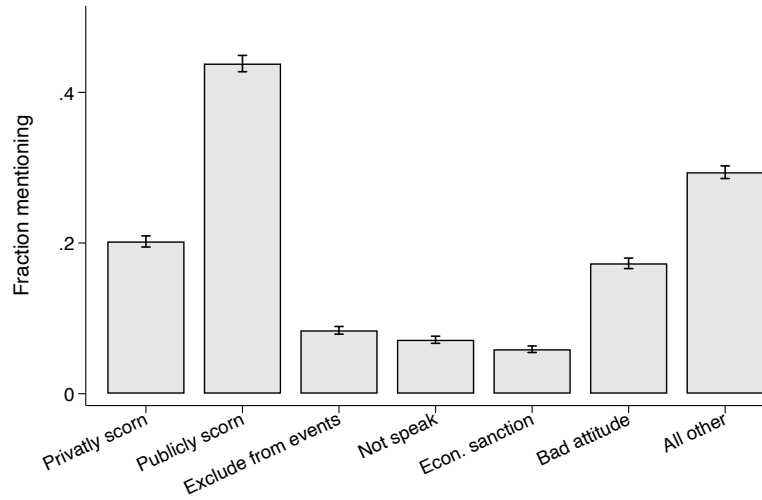
Panel B: Type of praise
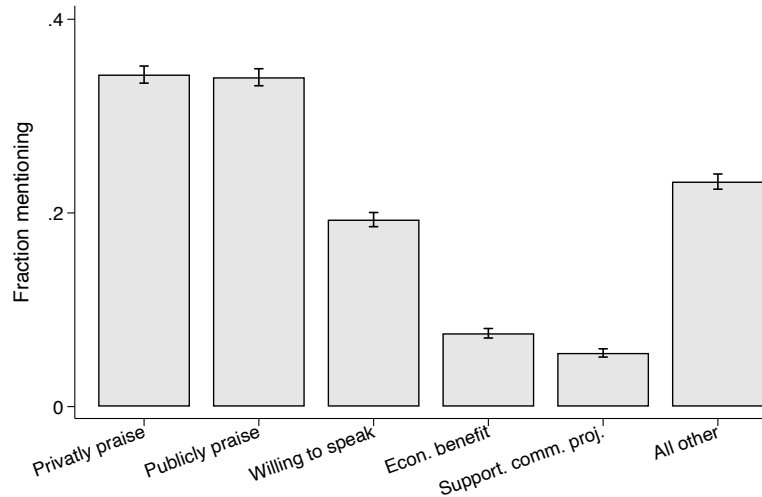
Notes: **Figure A.5 shows that women expect reactions to meeting attendance to consist mostly of social and not economic nature.** The figures display the share of women who select a given type of reaction in response to the question "*Imagine that villagers socially praised/punished you. What do you think? How might other villagers praise/punish you?*" Respondents could select multiple options. Economic sanctions include the following options: exclude from private or public resources, refuse to engage in business, refuse to lend money, and refuse to give presents. Economic benefits include the following options: share private resources, engage in business, lend money, and give presents. Other options mostly consist of "do not know" responses.

B Appendix Tables

Table A1: Balance Table

	Pooled	Low support		High support		
	(1)	(2)	(3)	(4)	(5)	(6)
	Pooled	No MK	MK	No MK	MK	p(balanced)

Panel A: Demographics

Age	44.19	44.55	44.20	43.96	44.03	0.66
Female	1.00	1.00	1.00	1.00	1.00	.
Number of children	3.75	3.77	3.79	3.70	3.74	0.46
Married	0.86	0.86	0.85	0.87	0.86	0.55
Kyrgyz	0.96	0.96	0.96	0.95	0.95	0.56
Village health committee member	0.15	0.15	0.15	0.16	0.14	0.51
Housewife	0.41	0.39	0.40	0.41	0.42	0.46
Working for money	0.35	0.36	0.35	0.37	0.34	0.52
Retired	0.18	0.19	0.19	0.18	0.18	0.58
Primary education	0.02	0.03	0.02	0.02	0.03	0.45
Secondary education	0.72	0.72	0.73	0.73	0.72	0.78
Tertiary education	0.25	0.25	0.25	0.24	0.26	0.89

Panel B: Beliefs about support for female political activism

Belief: % ave. support	55.99	55.97	55.96	56.06	55.99	1.00
Neighbors' belief: % ave. support	50.71	49.38	51.30	51.23	50.92	0.08
Belief: % women support	68.98	68.85	69.34	68.82	68.89	0.94
Neighbors' belief: % women support	60.93	59.73	61.73	61.09	61.12	0.32
Belief: % men support	43.03	43.05	42.56	43.35	43.17	0.81
Neighbors' belief: % men support	40.62	39.29	41.06	41.37	40.74	0.11

Panel C: Intentions and activism

Likelihood of pol. action in next six months (0-4 Likert scale)	2.94	2.93	2.93	2.92	2.97	0.56
Interest in community grant program (0-4 Likert scale)	2.86	2.86	2.84	2.87	2.87	0.81
Number of observations	5201	1294	1355	1262	1290	

Notes: **Table A1 show that the treatment groups are balanced on observables.** The p-value of a Wald test of joint significance in a multinomial logit regression of the treatment variable on all covariates is 0.7873. The number of observations varies because of refusals (belief variables) or technical problems in early survey versions (education variables).

Table A2: Summary Statistics by Village Health Committee Membership

	Representative		VHC members		Comparison	
	(1) Mean	(2) N	(3) Mean	(4) N	(5) Δ	(6) p(Δ=0)
Panel A: Demographics						
Age	43.48	4426	48.26	775	4.78	0.00
Female	1.00	4426	1.00	775	0.00	.
Number of children	3.69	4426	4.12	775	0.43	0.00
Married	0.86	4426	0.84	775	-0.02	0.19
Kyrgyz	0.95	4426	0.97	775	0.02	0.02
Village health committee member	0.00	4426	1.00	775	1.00	.
Housewife	0.44	4426	0.23	775	-0.21	0.00
Working for money	0.32	4426	0.55	775	0.22	0.00
Retired	0.18	4426	0.18	775	0.00	0.96
Primary education	0.03	4426	0.01	701	-0.02	0.00
Secondary education	0.73	4426	0.72	701	-0.01	0.71
Tertiary education	0.25	4426	0.27	701	0.03	0.14
Panel B: Beliefs about support for female political activism						
Belief: % ave. support	55.57	4369	58.43	767	2.86	0.00
Neighbors' belief: % ave. support	50.41	4086	52.44	726	2.04	0.01
Belief: % women support	68.81	4402	69.97	769	1.17	0.22
Neighbors' belief: % women support	60.81	4163	61.57	734	0.75	0.50
Belief: % men support	42.35	4374	46.88	769	4.53	0.00
Neighbors' belief: % men support	40.15	4122	43.25	736	3.10	0.00
Panel C: Intentions and activism						
Likelihood of pol. action in next six months (0-4 Likert scale)	2.87	4426	3.35	775	0.48	0.00
Interest in community grant program (0-4 Likert scale)	2.76	4426	3.39	775	0.63	0.00

Notes: **Table A2 shows that village health committee members are older, more economically active, and more optimistic about prevailing support for female political activism.** The number of observations varies because of refusals (belief variables) or technical problems in early survey versions (education variables).

Table A3: Additional Belief Effects of Providing Information about High Support

	Δ Second-order belief		Belief levels		
	(1) Male	(2) Female	(3) Average	(4) Male	(5) Female
High condition	0.464*** (0.094)	0.374*** (0.100)	0.440*** (0.076)	0.505*** (0.088)	0.383*** (0.098)
Mean low condition	-0.085	-0.520	5.296	4.218	6.362
Observations	2508	2523	2513	2517	2531

Notes: **Table A3 shows that the effect of expert opinions is robust to using different measures of beliefs.** *High condition* indicates the effect of receiving a high rather than a low expert opinion about the prevailing support for female political activism. Sample restricted to individuals without mutual knowledge treatment. Columns 1 and 2 show the effect on the difference between post- and pre-treatment beliefs about the prevailing support for female political activism among men and women separately. Columns 3 to 5 show the effects on post-treatment belief levels. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Additional Perceived Mutual Knowledge Effects on Third-Order Beliefs

| | Δ Third-order belief | | Third-order belief levels | | |
| | (1) | (2) | (3) | (4) | (5) |
	Male	Female	Pooled	Male	Female
High condition	0.363***	0.355***	0.507***	0.504***	0.486***
	(0.098)	(0.106)	(0.082)	(0.090)	(0.105)
Mutual knowledge	-0.190**	-0.140	0.013	-0.063	0.087
	(0.095)	(0.109)	(0.083)	(0.089)	(0.107)
High condition × Mutual knowledge	0.369***	0.132	0.075	0.231*	-0.054
	(0.134)	(0.152)	(0.116)	(0.128)	(0.148)
Mean low / mutal knowledge	0.130	-0.198	4.893	4.055	5.741
Observations	4787	4821	4841	4886	4898

Notes: **Table A4 shows that perceived mutual knowledge most affects third-order belief about male support for female political activism.** Third-order beliefs refer to women's beliefs about their neighbors' belief about societal support for female political participation. *High condition* indicates the effect of receiving a high rather than a low expert opinion about the prevailing support for female political activism. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. Columns 1 and 2 show the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism among men and women separately (third-order beliefs). Columns 3 to 5 show the effects on post-treatment belief levels of the same measure. Effects are estimated using the full sample of respondents. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Fully Interacted Treatment Effects

	Δ Second-order belief	Δ Third-order belief	Attended meeting
	(1)	(2)	(3)
High condition	0.436***	0.365***	-0.014
	(0.072)	(0.079)	(0.013)
Mutual knowledge	-0.051	-0.169**	-0.039***
	(0.072)	(0.081)	(0.013)
High condition × Mutual knowledge	0.148	0.246**	0.030*
	(0.101)	(0.111)	(0.018)
Mean low / mutal knowledge	-0.292	-0.029	0.193
Observations	5083	4740	5201

Notes: **Table A5 shows that perceived mutual knowledge increases the effect of providing information about social norms.** *High condition* indicates the effect of receiving a high rather than a low expert opinion about the prevailing support for female political activism. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. Columns 1 and 2 show the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism (third-order beliefs). Columns 3 and 4 show the effect on a dummy indicating attendance at a community meeting. Columns 1 and 4 show how the effectiveness of the second-order belief treatment varies with the creation of perceived mutual knowledge. Columns 2 and 5 show the effect of perceived mutual knowledge in the low-support condition. Columns 3 and 6 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Treatment Effects on Intention Measures

	Intention index			Intended attedance		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.043 (0.051)			0.020 (0.018)		
Mutual knowledge		-0.024 (0.037)	0.024 (0.037)		-0.016 (0.013)	0.004 (0.013)
Mean no mutal knowledge	-0.005	0.014	-0.025	0.878	0.887	0.869
Observations	5201	2649	2552	5201	2649	2552

Notes: **Table A6 shows that treatment effects on intentions are qualitatively aligned with behavioral outcomes but less precisely estimated.** Δ *High condition* indicates the difference of the second-order belief treatment effect between the mutual knowledge and no mutual knowledge condition. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. Columns 1 to 3 show the effect on a pre-specified index of intentions to engage in political actions. This index includes the intention to participate in budget meetings, to contact at least one local leader in the next 6 months, to attempt to access official documents related to local service delivery, to vote in the next local election, to stand as a candidate in next local election, to join local village health committee in the next three months, and to submit a proposal to the community grant program. It also includes the intention to attend the community grants meeting which is displayed separately in columns 4 to 6. Sample is restricted to villages that are in the treatment group in Esenaliev et al. (2024). Columns 4 to 6 show the effect on a dummy indicating that a submitted a community grant application five to nine months after treatment. Columns 1 and 4 show how the effectiveness of the second-order belief treatment varies with the creation of perceived mutual knowledge. Columns 2 and 5 show the effect of perceived mutual knowledge in the low-support condition. Columns 3 and 6 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Main effects by gender of source of social norm

	Δ Third-order belief			Attended meeting		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.118 (0.154)			0.012 (0.026)		
Δ High condition × Female norm	0.259 (0.223)			0.037 (0.037)		
Mutual knowledge		-0.093 (0.111)	0.028 (0.111)		-0.032* (0.018)	-0.024 (0.019)
Mutual knowledge × Female norm		-0.143 (0.163)	0.105 (0.157)		-0.008 (0.026)	0.021 (0.027)
Effect of female norm	0.377** (0.161)	-0.236** (0.120)	0.133 (0.111)	0.049* (0.026)	-0.040** (0.019)	-0.003 (0.019)
Mean no mutal knowledge	0.148	-0.115	0.366	0.185	0.176	0.173
Observations	4740	2410	2329	5201	2649	2552

Notes: **Table A7 shows that there is no significant heterogeneity in treatment effects by whether the provided information concerns men's or women's attitudes.** Δ *High condition* indicates the difference of the second-order belief treatment effect between the mutual knowledge and no mutual knowledge condition. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. *Female norm* indicates being provided an expert opinion about the attitudes of women as opposed to men. Columns 1 to 3 show the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism (averaged across gender). Columns 4 to 6 show the effect on a dummy indicating attendance at a community meeting. Columns 1 and 4 show how the effectiveness of the second-order belief treatment varies with the creation of perceived mutual knowledge. Columns 2 and 5 show the effect of perceived mutual knowledge in the low-support condition. Columns 3 and 6 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Treatment Effects on Longer-Term Outcomes

| | Fractions trainings attended | | | Sumitted grant application | | |
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.007 (0.019)			-0.009 (0.009)		
Mutual knowledge		-0.024* (0.014)	-0.019 (0.014)		-0.006 (0.007)	-0.018*** (0.007)
Mean no mutal knowledge	0.127	0.137	0.117	0.036	0.032	0.039
Observations	2591	1330	1261	5201	2649	2552

Notes: **Table A8 shows treatment effects on longer-term behavioral outcomes.** Δ *High condition* indicates the difference of the second-order belief treatment effect between the mutual knowledge and no mutual knowledge condition. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. Columns 1 to 3 show the effect on the fraction of training sessions attended in the five months after treatment. Sample is restricted to villages that are in the treatment group in Esenaliev et al. (2024). Columns 4 to 6 show the effect on a dummy indicating that a submitted a community grant application five to nine months after treatment. Columns 1 and 4 show how the effectiveness of the second-order belief treatment varies with the creation of perceived mutual knowledge. Columns 2 and 5 show the effect of perceived mutual knowledge in the low-support condition. Columns 3 and 6 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Balance Table for Vignette Experiments

	Pooled	Low support		High support		
	(1)	(2) No MK	(3) MK	(4) No MK	(5) MK	(6) p(balanced)
Panel A: Demographics						
Age	45.79	45.69	45.52	45.57	46.37	0.39
Female	1.00	1.00	1.00	1.00	1.00	.
Married	0.87	0.86	0.88	0.87	0.87	0.68
Wage employed	0.36	0.37	0.36	0.34	0.37	0.79
Self employed	0.11	0.10	0.12	0.12	0.10	0.65
Retired	0.14	0.15	0.10	0.15	0.17	0.04
Tertiary education	0.27	0.26	0.25	0.28	0.28	0.52
Secondary education	0.57	0.58	0.58	0.58	0.55	0.47
Panel B: Attitudes						
Support female pol. activism (1 to 5)	3.81	3.82	3.82	3.78	3.82	0.56
Generalized trust (1 to 10)	7.32	7.41	7.39	7.20	7.29	0.21
Consider friends for pol. action (1 to 4)	3.11	3.11	3.10	3.09	3.15	0.45
Consider others for pol. action (1 to 4)	2.89	2.90	2.84	2.87	2.94	0.15
Number of observations	3760	948	992	890	930	

Notes: **Table A9 shows that treatment groups in the vignette experiment are balanced on observables.** "MK" indicates the mutual knowledge equivalent treatment. Sample consists of all women taking part in the vignette survey 10 to 12 months after the initial experiment.

Table A10: Vignette Experiment: Reactions to Meeting Non-Attendance

	Praise non-attendance			Punish non-attendance		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	-0.019 (0.015)			-0.012 (0.015)		
Mutual knowledge		0.004 (0.010)	-0.010 (0.011)		-0.016 (0.010)	-0.029*** (0.011)
Mean no mutal knowledge	0.052	0.053	0.052	0.058	0.057	0.058
Observations	3760	1940	1820	3760	1940	1820

Notes: **Table A10 shows that reactions to meeting attendance are mostly insensitive to mutual knowledge.** The vignette experiments create variation that mimics the main experiment. Columns 1 to 3 show effects on a dummy indicating that individuals would praise other women for not attending a community meeting. Columns 4 to 6 show effects on a dummy indicating that individuals would punish other women for not attending a community meeting. Columns 1 and 4 show how the difference between supportive and opposing norms around female political activism differs by mutual knowledge. Columns 2 and 5 show the effect of mutual knowledge when opposing norms are prevalent. Columns 3 and 6 show the effect of mutual knowledge when supportive norms are prevalent. Control variables include age, a self-employment dummy, a wage employment dummy, a dummy for being married, a dummy for being a village health committee member, dummies for having secondary, and tertiary education, the standardized support for female political activism, and measures of network size and structure (# friends, # friend talk about politics, # friends whose opinion matter), baseline treatment status, past meeting attendance, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Analysis of Persuasion Channel

	Soc. norm. agree			Soc. norm. disagree			Attended meeting					
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High	(7) Pooled	(8) Low	(9) High	(10) Pooled	(11) Low	(12) High
Δ High condition	0.036 (0.033)			0.001 (0.020)			0.046* (0.027)			0.040 (0.027)		
Mutual knowledge		-0.051** (0.024)	-0.017 (0.024)		0.002 (0.015)	0.008 (0.015)		-0.066*** (0.020)	-0.026 (0.020)		-0.056*** (0.020)	-0.023 (0.020)
Soc. norm. agree										0.165*** (0.015)	0.187*** (0.023)	0.145*** (0.022)
Soc. norm. disagree										-0.014 (0.018)	-0.002 (0.033)	-0.031 (0.023)
Mean no mutual knowledge	0.700	0.723	0.678	0.080	0.074	0.086	0.289	0.312	0.265	0.289	0.312	0.265
Observations	2897	1465	1430	2897	1465	1430	2897	1465	1430	2897	1465	1430

Notes: **Table A11 shows that, while we observe some evidence of persuasion effects, they do not mediate the main treatment effects.** Δ *High condition* indicates the difference of the second-order belief treatment effect between the mutual knowledge and no mutual knowledge condition. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment in sample split regressions. The sample consists of women interviewed in both the baseline and the vignette surveys. During the vignette survey about 12 months after the treatment, we also ask individuals for the agreement with the prescriptive norm "Women in Kyrgyzstan should be more active in politics." on a five-point Likert scale (Disagree strongly; disagree; neither agree nor disagree; agree; Strongly agree). Columns 1 to 3 show the effect on a dummy indicating agreement with the statement. Columns 4 to 6 show the effect on a dummy indicating disagreement with the statement. Columns 7 to 12 show effects on our main outcome (attending the village meeting shortly after treatment). Columns 10 to 12 additionally control for post-treatment agreement with the social norm. Columns 1, 4, 7, and 10 show the estimated Δ *High condition*. Columns 2, 5, 8, and 11 show the effect of perceived mutual knowledge in the low-support condition. Columns 3, 6, 9, and 12 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Effects on gendered beliefs by gender of source of social norm

	Δ Third-order belief (male)			Δ Third-order belief (female)		
	(1) Pooled	(2) Low	(3) High	(4) Pooled	(5) Low	(6) High
Δ High condition	0.299 (0.183)			-0.104 (0.193)		
Δ High condition $\times$ Female norm	-0.254 (0.256)			0.511* (0.273)		
Mutual knowledge		-0.029 (0.131)	0.285** (0.134)		-0.138 (0.135)	-0.227 (0.142)
Mutual knowledge $\times$ Female norm		0.190 (0.182)	-0.114 (0.183)		-0.048 (0.196)	0.443** (0.195)
Effect of female norm	0.045 (0.180)	0.161 (0.129)	0.171 (0.128)	0.406** (0.196)	-0.186 (0.145)	0.215 (0.137)
Mean no mutal knowledge	0.160	-0.061	0.513	-0.324	-0.580	-0.125
Observations	5097	2596	2501	5131	2613	2518

Notes: **Table A12 shows that belief updating about gender-specific norms is broadly in line with the gender the provided information concerned.** Δ *High condition* indicates the difference of the second-order belief treatment effect between the mutual knowledge and no mutual knowledge condition. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. *Female norm* indicates being provided an expert opinion about the attitudes of women as opposed to men. Columns 1 to 3 show the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism among men. Columns 4 to 6 show the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism among women. Columns 1 and 4 show how the effectiveness of the second-order belief treatment varies with the creation of perceived mutual knowledge. Columns 2 and 5 show the effect of perceived mutual knowledge in the low-support condition. Columns 3 and 6 show the effect of perceived mutual knowledge in the high-support condition. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Social norm intervention effects by gender of source of social norm

	Δ Second-order belief			Second-order belief
	(1)	(2)	(3)	(4)
	average	male	female	
High condition	0.407***	0.444***	0.373***	0.003
	(0.104)	(0.136)	(0.141)	(0.019)
High condition	0.008	0.032	-0.007	-0.033
× Female norm	(0.147)	(0.189)	(0.199)	(0.027)
Female norm	-0.224**	-0.238*	-0.223	0.013
	(0.104)	(0.131)	(0.142)	(0.019)
High condition effect	0.415***	0.476***	0.366***	-0.030
with female norm	(0.103)	(0.131)	(0.141)	(0.019)
Mean no mutual knowledge	-0.074	0.160	-0.324	0.185
Observations	2502	2508	2523	2556

Notes: **Table A13 shows that there is no significant heterogeneity in the effect of the social norm treatment by whether the provided information concerns men's or women's attitudes.** *High condition* indicates that the respondent saw the high information about prevailing social norms. *Mutual knowledge* indicates the effect of receiving the mutual knowledge treatment. *Female norm* indicates being provided an expert opinion about the attitudes of women as opposed to men. Sample restricted to individuals without mutual knowledge treatment Column 1 shows the effect on the difference between post- and pre-treatment beliefs about neighbors' beliefs about support for female political activism (averaged across gender). Columns 2 show the effect on a dummy indicating attendance at a community meeting. Control variables include age, a self-employment dummy, a wage employment dummy, number of children, a dummy for being married, a dummy for being a village health committee member, dummies for having primary, secondary, and tertiary education, and the standardized self-reported likelihood of engaging politically over the next six months, and village fixed effects. Robust standard errors are given in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Theory Appendix

This sections proves the results presented in Section 2 and describes the framework in more detail. Figure 1 displays the structure of the theoretical framework. During the first stage, N women i decide whether to engage in a political action a_i. We initially model women's decisions as independent to simplify the exhibition of the framework but allow for strategic interactions between women later on. In the second stage, the community decides whether to react to women's activism based on their private opinion and their perceptions of social norms. Women anticipate the community's reaction, which links their higher-order beliefs about social norms to levels of activism.

C.1 Stage two decision making

We start by describing the community's reaction to i's action. As displayed in Figure 1, we assume that the community reacts only to political activism and not to its absence. This simplifies the exhibition without affecting the key message of the framework. We assume that the community reacts based on two variables: First, they consider their private opinion about the desirability of political activism s_j ($s_j = 1$ indicates support for political activism). Second, they consider the prevailing social norm s_- about the desirability of political activism. Again, $s_- = 1$ indicates support for political activism. We model the social norm as binary to match the active control design of our experiment, where we only distinguish between a high and low support condition. Importantly, we allow them to potentially misperceive the true social norm and assume that they act according to their belief about the social norm $\widehat{s_-}$.

Finally, we assume that the community only reacts to actions a_i if private opinion and perceived social norms align. Hence, j's behavior is fully described by Table A14. The community will punish women who choose a_i iff they privately oppose a_i and they believe that the prevailing social norm also opposes a_i. Conversely, they will praise the action a_i iff

Table A14: Second stage community decision

| | Perceived social norm $\widehat{s_-}$ | |
Private opinion (s_j)	Oppose (0)	Support (1)
Oppose (0)	Punish	No action
Support (1)	No action	Praise

Notes: **Table A14 shows j's reaction to i choosing a_i depending on their private opinion (s_j) and perceived social norms ($\widehat{s_-}$).**

they privately support a_i and they believe that the prevailing social norm also supports a_i. They do not react in any other case.

C.2   Stage one decision making

Women i in the community anticipate the community reaction for their decision whether to choose $a_i = 1$. Equation 4 displays the utility function for women i:

$$ E\left[U_w(a_i, s_i, \tilde{s}_j, \widehat{\tilde{s}_-})\right] = a_i \left(\gamma_i + s_i - \delta_1 E\left[punish_j | \tilde{s}_j, \widehat{\tilde{s}_-}\right] + \delta_2 E\left[praise_j | \tilde{s}_j, \widehat{\tilde{s}_-}\right] \right) \quad (4) $$

where γ_i is the private utility i gains from taking action a_i net of the non-social cost of engaging in the action. We assume that γ_i is ex-ante distributed with CDF $F(\gamma_i)$. s_i is i's private opinion about the desirability of a_i. $\tilde{s}_j$ is i's second-order belief, that is her belief about whether the community j privately endorses action a_i. $punish_j$ and $praise_j$ indicate the community's reaction. $\delta_1, \delta_2 > 0$ capture the associated utility with the community reactions. $\widehat{\tilde{s}_-}$ is i's belief about the community's belief about the social norm ($\widehat{s_-}$), a proxy for the third-order beliefs we shift experimentally. We model beliefs about social norms as binary to match the structure of our experiment and to clarify the definition of perceived mutual knowledge as $\tilde{s}_j = \widehat{\tilde{s}_-}$

i chooses $a_i = 1$ iff $E\left[U_w(1, s_i, \tilde{s}, \widehat{\tilde{s}_-})\right] > E\left[U_w(0, s_i, \tilde{s}, \widehat{\tilde{s}_-})\right]$. This yields the following

condition:

$$\gamma_i > -s_i + \delta_1 E\left[punish_j | \tilde{s}, \widehat{\tilde{s}_-}\right] - \delta_2 E\left[praise_j | \tilde{s}, \widehat{\tilde{s}_-}\right] \tag{5}$$

This yields four distinct thresholds of $\gamma^*(\tilde{s}, \widehat{\tilde{s}_-})$ above which i engages in in action a_i:

1. If $\tilde{s}_j = 1$ and $\widehat{\tilde{s}_-} = 1$: $\gamma^*_{1,1} = -s_i - \delta_2$

2. If $\tilde{s}_j = 0$ and $\widehat{\tilde{s}_-} = 1$: $\gamma^*_{0,1} = -s_i$

3. If $\tilde{s}_j = 1$ and $\widehat{\tilde{s}_-} = 0$: $\gamma^*_{1,0} = -s_i$

4. If $\tilde{s}_j = 0$ and $\widehat{\tilde{s}_-} = 0$: $\gamma^*_{0,0} = -s_i + \delta_1$

The likelihood that i chooses $a_i = 1$ is $P(a_i = 1) = 1 - F(\gamma^*(\tilde{s}, \widehat{\tilde{s}_-}))$.

Proof of result 1: To assess whether divergence in second-order beliefs induce a behavioral wedge, we define the participation gap between the cases supportive and opposing social norms, conditional on third-order beliefs:

$$\Delta P(\widehat{\tilde{s}_-}) = 1 - F(\gamma^*(1, \widehat{\tilde{s}_-}) - (1 - F(\gamma^*(0, \widehat{\tilde{s}_-})) = F(\gamma^*(0, \widehat{\tilde{s}_-})) - F(\gamma^*(1, \widehat{\tilde{s}_-}) > 0 \tag{6}$$

The inequality holds for both $\widehat{\tilde{s}_-} = 1$ and $\widehat{\tilde{s}_-} = 0$ as $\gamma^*(1, 0) = -s_i < \gamma^*(0, 0) = -s_i + \delta_1$ and $\gamma^*(1, 1) = -s_i - \delta_2 < \gamma^*(0, 1) = -s_i$. Hence, there will be higher levels of political activism in the case of perceived social support relative to the case of perceived opposition.

Proof of result 2: To prove result 2, we consider the participation gap for the case of mutual knowledge. This yields:

22

$$\Delta P_{mut} = F(\gamma^*(0,0)) - F(\gamma^*(1,1)) = F(-s_i + \delta_1) - F(-s_i - \delta_2) > 0 \tag{7}$$

As F is strictly increasing, we know that $F(-s_i + \delta_1) - F(s_i - \delta_2) > F(-s_i + \delta_1) - F(s_i)$ and $F(-s_i + \delta_1) - F(s_i - \delta_2) > F(-s_i) - F(s_i - \delta_2)$. This directly implies that $\Delta P_{mut} > \Delta P(\widehat{\tilde{s}_-}) \ \forall \ \widehat{\tilde{s}_-}$. Hence, mutual knowledge increases the behavioral wedge between supportive and opposing social norms as stated in result 2.

C.3 Strategic interactions

We incorporate strategic interactions by letting the private returns to acting vary with the fraction of other women who act. That is we define $\gamma_i = \phi^k(\bar{a}) + \alpha_i$. α_i is distributed with CDF G. $\phi^k(\bar{a})$ captures strategic interactions by making the return to political activism a function of the fraction of other women who choose $a_i = 1$ ($\bar{a} = \frac{1}{N} \sum_i^N a_i$). Where $k \in \{comp, none, sub\}$ indicates the type of strategic interaction. $\frac{d\phi^{comp}(\bar{a})}{d\bar{a}} > 0$ implies strategic complementarity while $\frac{d\phi^{sub}(\bar{a})}{d\bar{a}} < 0$ implies strategic substitutability. $\frac{d\phi^{none}(\bar{a})}{d\bar{a}} = 0$ is the absence of strategic interactions. We make one further normalization assumption to make the ϕ^k comparable in levels. That is we assume that $\phi^k(1 - F(-1 + \delta_1) + \epsilon) = 0 \ \forall k$ and ϵ close to zero. This is purely a normalization stating that ϕ^k is zero at some point between $\bar{a}(1,0)$ and $\bar{a}(0,1)$ (see below). It is ex-ante unclear which of the two patterns we would expect, as the existing literature documents mixed results (e.g., Cantoni et al., 2019; González, 2020; Hager et al., 2022).

We can now define a threshold for α_i that determines the participation decision.

This yields four distinct thresholds of $\alpha^*(\tilde{s}, \widehat{\tilde{s}_-})$ above which i engages in in action a_i:

1. $\alpha_{1,1}^* = -s_i - \phi(\bar{a}(1,1)) - \delta_2$

2. $\alpha_{0,1}^* = -s_i - \phi(\bar{a}(0,1))$

3. $\alpha^*_{1,0} = - - s_i - \phi(\bar{a}(1,0))$

4. $\alpha^*_{0,0} = -s_i - \phi(\bar{a}(0,0)) + \delta_1$

For simplicity, we assume that $\tilde{s}_j = 1$ implies that i believes that all other women l in the community support the action (i.e., $s_l = 1$) and $\tilde{s}_j = 0$ implies that i believes that all other women l in the community oppose the action (i.e., $s_l = 0$). We also assume that $\widehat{\tilde{s}_-} = 1$ implies that other women perceive private support of others. We also have to make assumptions about beliefs about other women's third-order beliefs. To limit the complexity to the scope of the experiment, we assume that these beliefs always match third-order beliefs $\widehat{\tilde{s}_-}$. Finally, we assume that N is large enough so that $P(a_i = 1) \approx \bar{a}$.

For now, we assume that women do not anticipate the strategic responses of others and that they act according to the framework without strategic interactions. We can then define the anticipated fraction of women l who participate as $\bar{a}$ as:

$$\bar{a}(\tilde{s}_j, \widehat{\tilde{s}_-}) = 1 - F\left(-\tilde{s}_j - \delta_2 \widehat{\tilde{s}_-}^2 + \delta_1(1 - \widehat{\tilde{s}_-})^2\right) \tag{8}$$

From this, we can compare the average levels of anticipated participation across beliefs and obtain the following ranking. For a complete ordering, we make the additional assumption that the fear of punishment cannot be so strong as to offset the combined effect of a supportive private opinion and anticipated praise, that is $\delta_2 < \delta_1 - 1$.

$$\bar{a}(1, 1) = 1 - F(-1 - \delta_2) \quad > \quad \bar{a}(1, 0) = 1 - F(-\delta_2) > \tag{9}$$
$$\bar{a}(0, 1) = 1 - F(-1 + \delta_1) \quad > \quad \bar{a}(0, 0) = 1 - F(\delta_1)$$

We can then plug this into i's decision problem and obtain the following threshold for α^*:

$$\alpha^*(\tilde{s}, \widehat{\tilde{s}_-}) = -s_i - \phi^k\left(\bar{a}(\tilde{s}_j, \widehat{\tilde{s}_-})\right) - \delta_2 \tilde{s}\widehat{\tilde{s}_-} + \delta_1(1-\tilde{s})(1-\widehat{\tilde{s}_-}) \tag{10}$$

Under complementarity we get $\phi^{comp}\left(\bar{a}(1, \widehat{\tilde{s}_-})\right) > \phi^{comp}\left(\bar{a}(0, \widehat{\tilde{s}_-})\right)$. Under strategic substitutability we obtain the opposite $\phi^{sub}\left(\bar{a}_h\right) < \phi^{sub}\left(\bar{a}_l\right)$.

With the normalization of $\phi^k\left(1 - F\left(-1 + \delta_1\right) + \epsilon\right) = 0 \;\; \forall k$, we get two further inequalities: $\phi^{comp}\left(\bar{a}(1, \widehat{\tilde{s}_-})\right) > 0 > \phi^{sub}\left(\bar{a}(1, \widehat{\tilde{s}_-})\right)$ and $\phi^{comp}\left(\bar{a}(0, \widehat{\tilde{s}_-})\right) < 0 < \phi^{sub}\left(\bar{a}(0, \widehat{\tilde{s}_-})\right)$.

We can now compare the participation gap for different sets of beliefs about social norms. First, consider the gap for fixed third-order beliefs $\widehat{\tilde{s}_-}$:

$$\begin{aligned}
\Delta P^k(\widehat{\tilde{s}_-}) =\;\; & G\left[-s_i - \phi^k\left(\bar{a}(0, \widehat{\tilde{s}_-})\right) + \delta_1(1 - \widehat{\tilde{s}_-})\right] \\
& - \;\; G\left[-s_i - \phi^k\left(\bar{a}(1, \widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right]
\end{aligned} \tag{11}$$

With the assumption we can obtain result 3:

$$\Delta P_s^{comp}(\widehat{\tilde{s}_-}) > \Delta P_s^{none}(\widehat{\tilde{s}_-}) > \Delta P_s^{sub}(\widehat{\tilde{s}_-}) \tag{12}$$

We prove each inequality in turn:

$$\Delta P^{comp}(\widehat{s_-}) > \Delta P^{none}(\widehat{s_-})$$

$$\Rightarrow \; G\left[-s_i - \phi^{comp}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i - \phi^{comp}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right] >$$
$$G\left[-s_i + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i - \delta_2\widehat{\tilde{s}_-}\right]$$

$$\Rightarrow \quad G\left[-s_i - \phi^{comp}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i + \delta_1(1-\widehat{\tilde{s}_-})\right] >$$
$$G\left[-s_i - \phi^{comp}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right] - G\left[-s_i - \delta_2\widehat{\tilde{s}_-}\right]$$

$$\Rightarrow \quad G\left[-s_i - \phi^{comp}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i + \delta_1(1-\widehat{\tilde{s}_-})\right] > 0 >$$
$$G\left[-s_i - \phi^{comp}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right] - G\left[-s_i - \delta_2\widehat{\tilde{s}_-}\right]$$

For the last step note that the left-hand side is greater 0 because $\phi^{comp}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) < 0$. The right-hand side is smaller zero because $\phi^{comp}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) > 0$.

For the second inequality, consider the following:

$$\Delta P^{sub}_s(\widehat{s_-}) < \Delta P^{none}_s(\widehat{s_-})$$

$$\Rightarrow \; G\left[-s_i - \phi^{sub}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i - \phi^{sub}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right] <$$
$$G\left[-s_i + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i - \delta_2\widehat{\tilde{s}_-}\right]$$

$$\Rightarrow \quad G\left[-s_i - \phi^{sub}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i + \delta_1(1-\widehat{\tilde{s}_-})\right] <$$
$$G\left[-s_i - \phi^{sub}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right] - G\left[-s_i - \delta_2\widehat{\tilde{s}_-}\right]$$

$$\Rightarrow \quad G\left[-s_i - \phi^{sub}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i + \delta_1(1-\widehat{\tilde{s}_-})\right] < 0 <$$
$$G\left[-s_i - \phi^{sub}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right] - G\left[-s_i - \delta_2\widehat{\tilde{s}_-}\right]$$

For the last step again note that the left-hand side is smaller 0 because $\phi^{sub}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) > 0$. The right-hand side is greater zero because $\phi^{sub}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) < 0$.

Mutual knowledge Next, we show that the effect of mutual knowledge on the behavioral gap is amplified under strategic complementarity. Formally, we proof that:

$$\Delta P_{mut}^{comp} > \Delta P^{comp}(\widehat{\tilde{s}_-}) > \Delta P^{sub}(\widehat{\tilde{s}_-}) > \Delta P_{mut}^{sub} \tag{13}$$

We proof this in parallel to the previous subsection. First, consider the case of strategic complementarity:

$$\Delta P_{mut}^{comp} > \Delta P^{comp}(\widehat{\tilde{s}_-})$$

$$\Rightarrow \quad G\left[-s_i - \phi^{comp}\left(\bar{a}(0,0)\right) + \delta_1\right] - G\left[-s_i - \phi^{comp}\left(\bar{a}(1,1)\right) - \delta_2\right] >$$

$$G\left[-s_i - \phi^{comp}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] - G\left[-s_i - \phi^{comp}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right]$$

$$\Rightarrow \quad G\left[-s_i - \phi^{comp}\left(\bar{a}(0,0)\right) + \delta_1\right] - G\left[-s_i - \phi^{comp}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1-\widehat{\tilde{s}_-})\right] >$$

$$G\left[-s_i - \phi^{comp}\left(\bar{a}(1,1)\right) - \delta_2\right] - G\left[-s_i - \phi^{comp}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right]$$

We can now consider two cases to prove that this inequality holds. First, for $\widehat{\tilde{s}_-} = 0$, the left-hand side of the inequality is equal to zero and the right-hand side is smaller than zero because $-\phi^{comp}\left(\bar{a}(1,1)\right) - \delta_2 < -\phi^{comp}\left(\bar{a}(1,0)\right)$. Second, for $\widehat{\tilde{s}_-} = 1$, the left-hand side of the inequality is greater zero because $-\phi^{comp}\left(\bar{a}(0,0)\right) + \delta_1 > -\phi^{comp}\left(\bar{a}(0,1)\right)$ and the right-hand side is equal to zero.

Finally, we show that the effect of strategic substitutability on the effect of mutual knowledge on the behavior gap is negative only if the strategic interactions are sufficiently strong:

$$\Delta P_{mut}^{sub} < \Delta P^{sub}(\widehat{\tilde{s}_-})$$

$$\Rightarrow \quad G\left[-s_i - \phi^{sub}\left(\bar{a}(0,0)\right) + \delta_1\right] - G\left[-s_i - \phi^{sub}\left(\bar{a}(1,1)\right) - \delta_2\right] <$$

$$G\left[-s_i - \phi^{sub}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1 - \widehat{\tilde{s}_-})\right] - G\left[-s_i - \phi^{sub}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right]$$

$$\Rightarrow \quad G\left[-s_i - \phi^{sub}\left(\bar{a}(0,0)\right) + \delta_1\right] - G\left[-s_i - \phi^{sub}\left(\bar{a}(0,\widehat{\tilde{s}_-})\right) + \delta_1(1 - \widehat{\tilde{s}_-})\right] <$$

$$G\left[-s_i - \phi^{sub}\left(\bar{a}(1,1)\right) - \delta_2\right] - G\left[-s_i - \phi^{sub}\left(\bar{a}(1,\widehat{\tilde{s}_-})\right) - \delta_2\widehat{\tilde{s}_-}\right]$$

Again, consider two cases for $\widehat{s_-}$ to derive the conditions under which the inequality holds. First, for $\widehat{s_-} = 0$, the left-hand side of the inequality is equal to zero and the right-hand side is smaller than zero if $\phi^{sub}\left(\bar{a}(1,0)\right) - \phi^{sub}\left(\bar{a}(1,1)\right) > \delta_2$. Second, for $\widehat{s_-} = 1$, the right-hand side is equal to zero and the left-hand side of the inequality is greater than zero if $\delta_1 > \phi^{sub}\left(\bar{a}(0,0)\right) - \phi^{sub}\left(\bar{a}(0,1)\right)$. Only if these two conditions hold does strategic substitutability necessarily imply a negative effect of mutual knowledge on the behavior gap. If strategic interactions are sufficiently weak, mutual knowledge may increase the behavior gap even under strategic substitutability.

D   Deviations from the pre-analysis plan

This sections describes how we deviated from the pre-analysis plan for this study (`https://aspredicted.org/dqm5-drwg.pdf`). Our main analysis largely sticks to the pre-analysis plan. We only make the following changes:

1. We study effects on changes in beliefs instead of belief levels in the main text to increase statistical power (pre-specified outcomes are report in Table A4).

2. We focus on meeting attendance as the most short-term outcome as information spillovers likely attenuate medium to long-term effects of the treatment. We report

the pre-registered medium term outcome (grant application submission) and the additional training participation outcome in Table A8.

3. We do not study effects on pre-registered endline participation outcomes. For budget reasons we could not re-interview all women who took part the initial experiment.

E   Items on planned political participation

1. Now we will ask you questions about the support of women in participation in local politics "To what extent do you agree with the following statements?"

   - Generally speaking, men in my village would support it if women like me participate more in local politics.

   - Generally speaking, my neighbors think that men in my village would support it if women like me participate more in local politics.

   - Generally speaking, women in my village would support it if women like me participate more in local politics.

   - Generally speaking, my neighbors think that women in my village would support it if women like me participate more in local politics.

   - My husband/partner would support it if women like me participate more in local politics.

2. Over the next six months, how likely is it that you will participate in meetings organized by local municipality or citizens in your area that citizens could attend? This includes local budget meetings or Tuloo.

3. Imagine that you were to attend such a meeting over the next 6 six months. How likely would you be to speak in such a meeting?

4. How likely is it for you to contact the following political leaders?

- Local government councilor

- Member of parliament

- Official of a government agency (Ministries, social fund, tax, police, state register, etc.)

- Traditional leader (Court of Elders, clan leaders etc.)

- Political party official

- Religious leader

- Civil society leader

5. Over the next six months, how likely is it that you will ask anyone about the budget project of local municipalities, i.e. how they are distributed and which sector to improve?

6. Over the next six months, how likely is it that you will try to access any documents related to local services delivered by local municipality (e.g. transportation, greening, education, health etc)?

7. How likely are you to vote in the next local election?

8. How likely are you to stand as a candidate in the next local election?

9. Think of the next three months. How likely are you to join the village health committee?