Experimental Economics (2020) 23:392–420
https://doi.org/10.1007/s10683-019-09616-z

ORIGINAL PAPER

'Everybody's doing it': on the persistence of bad social norms

David Smerdon[1] · **Theo Offerman**[2] · **Uri Gneezy**[3]

Received: 5 July 2017 / Revised: 15 May 2019 / Accepted: 21 May 2019 / Published online: 31 May 2019

Abstract
We investigate how information about the preferences of others affects the persistence of 'bad' social norms. One view is that bad norms thrive even when people are informed of the preferences of others, since the bad norm is an equilibrium of a coordination game. The other view is based on pluralistic ignorance, in which uncertainty about others' preferences is crucial. In an experiment, we find clear support for the pluralistic ignorance perspective . In addition, the strength of social interactions is important for a bad norm to persist. These findings help in understanding the causes of such bad norms, and in designing interventions to change them.

Keywords Social norms · Pluralistic ignorance · Social interactions · Equilibrium selection · Conformity

JEL Classification C92 · D70 · D90 · Z10

1 Introduction

Social norms provide informal rules that govern our actions within different groups and societies and across all manner of situations. Many social norms develop in order to overcome market failure, mitigate negative externalities or promote positive

We acknowledge the University of Amsterdam Behavior Priority Area for providing funding for the experiment.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10683-019-09616-z) contains supplementary material, which is available to authorized users.

✉ David Smerdon
d.smerdon@uq.edu.au

[1] University of Queensland, Brisbane, Australia

[2] University of Amsterdam, Amsterdam, Netherlands

[3] Rady School of Management, UCSD, San Diego, USA

ones so as to facilitate some collective goal (Arrow 1970; Hechter and Opp 2001). However, social norms that are inefficient from a welfare perspective also persist.

A key feature of a social norm is the desire to conform to the majority in a group. We follow Bicchieri (2017, p. 35), who defines a social norm as a "rule of behavior such that individuals prefer to conform to it on condition that they believe that (a) most people in their relevant network conform to it (empirical expectation), and (b) most people in their relevant network believe they ought to conform to it (normative expectation) and may sanction deviations." We define a 'good' social norm as a norm that is supported in an equilibrium that maximizes group welfare, and a 'bad' social norm as a norm that is supported in an equilibrium that does not.

Sometimes good norms become bad norms when over time the payoff structure changes such that the norm ceases to be good for the group. One such example is provided by norms of revenge. In societies where there is no or minimal rule of law to enforce individual rights, tit-for-tat strategies are more likely to emerge as substitutes so that people can defend their property and honor (Elster 1990). In such settings where people cannot rely on the protection of an authority, norms involving revenge behavior can be welfare-enhancing (Elster 1989). The danger of such norms is that if someone transgresses—and, for instance, kills a neighbor in a fight—this may trigger a long, bloody encounter between families. A culture of these so-called 'blood feuds' can persist even after political or legal transitions have eroded the social benefits, or after a group has migrated to an area with an effective government and legal system (İçli 1994; Grutzpalk 2002). In either case, the persistence of the social norm no longer serves its purpose and becomes welfare-damaging.

Another example is that of gender roles in the labor market. Alesina et al. (2013) find causal evidence that traditional agricultural practices explain differences in attitudes towards workforce participation today, and that this effect is driven by persistent cultural norms.[1] Specialization of production along gender lines may have been a good equilibrium under certain conditions, such as when particularly productive occupations were physically demanding, home production was less efficient, and/or women had a comparative advantage in mentally-intensive tasks (Galor and Weil 1996; Greenwood et al. 2005). The traditional gender roles and cultural beliefs stemming from these labor conditions are supported by social sanctions for deviations and can persist even when the economy moves away from such an environment, or when groups migrate to a more developed country (Fernandez and Fogli 2009; Jayachandran 2015).

In this paper, we investigate conditions under which bad norms persist. We experimentally test situations in which a bad norm initially emerges as a good norm, but changes to the environment over time alter the payoff structure such that the norm ceases to solve negative externalities and actually begins to promote them. The most important contribution of our paper is that we investigate the extent to which a lack of information about others' preferences or attitudes is important for the development of bad social norms. In particular, we provide evidence that compares two

[1] The arguments for the negative welfare effects of restricting female labor participation on growth have been long discussed in economics; e.g. Goldin (1986).

predominant but opposing views on how information about others' preferences shapes bad norms.

One perspective is that bad norms can thrive independent of whether or not people are informed of the preferences of others. This view is supported by the model of Brock and Durlauf (2001), whose approach we use to study the development of norms. They propose a stationary coordination game in which agents are driven by a taste for conformity. All other things equal, agents benefit when more people make the same choice as they do. In equilibrium, players either coordinate on the welfare-maximizing allocation or on a welfare-inefficient allocation. We consider different versions of their game by allowing players to be uncertain about the preferences of others, and by allowing preferences to change over time.[2]

The alternative view is that bad norms are driven by pluralistic ignorance. Pluralistic ignorance refers to a situation in which most individuals have private attitudes and judgments that differ from the prevailing norm, and wrongly believe that the majority of group members have a private preference to keep to the status quo (Miller and McFarland 1987; Katz and Allport 1931). As a result, a bad norm may persist even though the majority of the group would like to change it.[3] Notice that in this approach, the uncertainty about other individuals' preferences is a necessary condition (Sherif 1936). In combination with wrong beliefs about the preferences of others, it may lead to the emergence and persistence of bad social norms.

We use the model of Brock and Durlauf (2001) to design an experiment that allows us to investigate the role that information about others' preferences plays for the development of bad norms, thus incorporating the insights from the pluralistic ignorance viewpoint. In our experiment, we monetarize social payoffs.[4] In a setup with a relatively small group size and in which the benefits from coordinating are large (with a strong 'social value' component), a setup that we conjecture to be particularly conducive for bad social norms, we vary the information about others' preferences in two ways. First, we compare a version of the game in which subjects

[2] Recent alternative approaches to modeling social norms include Michaeli and Spiro (2017), who focus on pairwise interactions in a coordination game, and Acemoglu and Jackson (2015), who investigate an intergenerational context.

[3] Pluralistic ignorance has been linked to the propagation of various damaging social issues, such as college binge-drinking (Prentice and Miller 1996; Schroeder and Prentice 1998), tax avoidance (Wenzel 2005), school bullying (Sandstrom et al. 2013), the spread of HIV/AIDS due to stigmas against condom usage (Gage 1998) and the lack of female labor force participation in Saudi Arabia (Bursztyn et al. 2018).

[4] We think that our results also shed light on situations where utility is derived from conforming to one's group identity instead of from a material payoff (Akerlof and Kranton 2000; Tajfel and Turner 1986, 1979). A raft of recent empirical evidence has demonstrated that social identity can influence individual decision-making and behavior in a wide range of respects, such as group problem-solving (Chen and Chen 2011), polarization of beliefs (Hart and Nisbet 2011; Luhan et al. 2009), preferences over outcomes (Charness et al. 2007), trust (Hargreaves Heap and Zizzo 2009), redistribution preferences (Chen and Li 2009), punishment behavior (Abbink et al. 2010), discrimination (Fershtman and Gneezy 2001), self-control (Inzlicht and Kang 2010), competitiveness (Gneezy et al. 2009) and time horizons for decision-making (Mannix and Loewenstein 1994). Several studies have successfully induced group identity directly in the lab to test for different effects; e.g., Chen and Chen (2011), Charness et al. (2007), Eckel et al. (2007), among others.

are uncertain about others' preferences with a version in which subjects are fully informed. Second, in the version in which they are uncertain about others' preferences, we vary whether subjects can communicate about their intended actions.

We use the version of the game with incomplete information about others' preferences to investigate some other variables that may affect how likely it is that people adhere to norms after they have turned to bad norms: (i) the social value component and (ii) group size. In fact, a strong social value component is essential for the existence of bad social norms in the approach of Brock and Durlauf (2001).

The experimental results show that the information about others' preferences is crucial for the emergence and persistence of bad norms. In agreement with pluralistic ignorance, when subjects are fully informed about others' preferences, groups move swiftly away from a good equilibrium after it has become bad. Allowing subjects to communicate also helps to escape norms that have become bad. Communication reduces the uncertainty about other subjects' preferences and intended behavior. On the other hand, bad norms thrive when subjects are uncertain about the payoffs of others.

While the main results support the pluralistic ignorance perspective, other experimental results within an incomplete information environment accord well with the Brock and Durlauf approach. In particular, the stronger the social value component, the more likely a bad norm is to persist. This result resonates with the finding in minimum effort games that it is more difficult to coordinate on the 'good' equilibrium when it is individually more costly to do so (for example, Devetag and Ortmann 2007). We find that bad norms are more likely to persist in larger groups in the short term, but this effect disappears in the long run.

Our paper contributes to a literature on the emergence and persistence of bad norms. Devetag and Ortmann (2007) survey how bad outcomes can emerge in team production processes that are characterized by a minimum effort production function. In the minimum effort game, players simultaneously exert costly effort, and the minimum effort in the team determines its productivity. The stage game hosts a multitude of Pareto-ranked equilibria. In agreement with risk dominance, subjects in experiments usually quickly coordinate on a bad equilibrium that offers them a secure but low payoff, unless group size is very small (Van Huyck et al.1990; Knez and Camerer 1994).[5] A special feature of the minimum effort game is that if only one 'rotten apple' provides low effort, all other players want to choose the same low effort. In the framework of Bicchieri (2017, p. 35), the minimum effort game is not about norms, because it is not a game where players want to follow what the majority in the group does (except in the cases where the majority choice coincides with the minimum). In addition, in the game that we study in the experiment the good equilibrium always risk-dominates the bad equilibrium, so risk dominance by itself cannot explain the persistence of bad norms that we observe in some circumstances.

[5] It appears to be very hard to avoid bad outcomes in minimum effort games, but there are some reliable factors that help subjects coordinate on better outcomes (Anderson et al. 2001; Brandts and Cooper 2006, 2007; Cachon and Camerer 1996; Cason et al. 2012; Chaudhuri et al. 2009; Chen and Chen 2011; Kopanyi-Peuker et al.2015; Riedl et al. 2015; Weber 2006).

In a similar vein, Lim and Neary's (2016) experimental investigation of stochastic adjustment dynamics also uses a large binary-action population game, the language game, in which individuals' choices are strategic complements. They find strong evidence that individuals behave consistently with a best-response learning rule based on the previous period's outcomes, which, in a noisy environment, can lead to groups escaping coordination equilibria. While our results are broadly consistent with this literature, our game also yields different insights: in the game that we study incomplete information on others' preferences is needed for the persistence of bad norms, while in the minimum effort and language games the bad outcome results even with complete information about preferences.

More recently, Abbink et al. (2017) identify an alternative driver of bad norms. The central insight from their experiment is that punishment opportunities can, under certain circumstances, lead to socially destructive norms being enforced in public good games. Specifically, in a linear public good game where group members only marginally benefit from others' contributions, such that the socially optimal act is to not contribute, they find that subjects support a bad social norm when they have the possibility to punish free-riders. The key difference between their approach and ours is that they study the emergence of bad social norms in inefficient public good provision, whereas we focus on pure coordination situations in which the question is whether groups can move from one equilibrium to a better one.

Closely related to our paper in terms of experimental design are Andreoni et al. (2017) and Duffy and Lafky (2018). Andreoni et al.'s (2017) investigation of so-called 'conformity traps', conceived independently and concurrently, complements our approach. The most important differences in design are the information environment and the payoffs pertaining to individuals who deviate from a norm. Individuals in their experiment know the true distribution and evolution of group preferences, such that pluralistic ignorance cannot play a role. By comparison, in our setup the individuals who deviate first from the current norm incur disproportionately large costs for pioneering the change, creating stronger incentives to wait for others to deviate first.[6] A particularly relevant feature of their results is that bad norms can still persist with full information over group preferences, so long as the strength of social payoffs is sufficiently high. Other than that, their main results are consistent with our own: (1) The scale of social payoffs, relative to individualistic utility differences, is crucial for conformity to a bad equilibrium, (2) Smaller groups can break a conformity trap faster, and (3) Anonymous communication through polls can aid escaping a conformity trap.

Duffy and Lafky (2018) independently study the role of the strength of conformity and uncertainty about others' preferences to test under which circumstances people choose against their privately held preferences. There are a couple of differences with our study. First, their model looks at two types rather than a continuum of individual preferences. Second, in their full information condition, subjects know the progression of preferences. Third, in their incomplete information condition,

[6] Their design also differs in terms of the matching structure: in each round, matches are pairwise with external payoffs for group conformity, rather than group coordination.

subjects know that in each period with probability 0.75 one player's type will switch. Despite these and other minor differences, their results agree qualitatively with our results.

The remainder of the paper is organized as follows. Section 2 presents the game and some theoretical benchmarks. In Sect. 3, we detail the design and procedure used to transpose the model into the laboratory, and we present the equilibrium predictions for the experimental parameters. Section 4 discusses the experimental results, from which insights into the factors affecting bad norm persistence are presented, and Sect. 5 concludes with a discussion of the results and implications.

2 Game and theoretical benchmarks

We adopt Brock and Durlauf's (2001) model of discrete choice with social interactions, with minor modifications, as a vehicle for investigating the persistence of bad norms in an experiment.

2.1 The stage game with full information on the common values

N players choose between two options. For example, teenagers in a social group decide whether or not to smoke. A player's payoff from the chosen option is composed of her *private value* and her *social value*, which measures the congruence between the player's choice and those of the group. Every player knows that each option's private value is comprised of the sum of a *common value* and a player-specific *private shock*. The (continuous) distributions generating the private shocks for the two choices are known to all the players. In the full information stage game, players are informed of the common values for each choice, and of their own private shocks for each choice (but not of the private shocks of the others) at the start of the stage game. In line with the approach of Brock and Durlauf (2001), player i receives a payoff of:

$$V_i(\omega_i) = u_{\omega_i} + S(\omega_i, \omega_{\neg i}) + \epsilon_i(\omega_i), \qquad \omega_i \in \{-1, 1\} \tag{1}$$

Here ω represents the choice variable, taking the value of 1 (smoking) or -1 (not smoking); u_{ω_i} represents the common value that pertains to player i*s* specific choice ω_i and is the same for any player choosing $\omega_j = \omega_i$; and $\epsilon_i(\omega_i)$ is a player choice-dependent shock. The shocks $\epsilon_i(\omega_i)$ have a mean of 0 and are identically and independently distributed across all players and choices such that the difference $\epsilon_i(-1) - \epsilon_i(1)$ has a known probability distribution function $F(\cdot)$.

$S(\omega_i, \omega_{\neg i})$ gives the social value of the choice that depends on player i's choice ω_i and the choices of all other players $\omega_{\neg i}$. In this game, the assumption is made that the utility derived from social payoffs exhibits "constant and totalistic strategic complementarity" (Brock and Durlauf 2001, p. 238), which is also employed in Andreoni et al. (2017)'s design. This means that players are always happier by the same amount when one more person makes the same choice as them. With this assumption, the form of social value is stipulated in (2):

$$S(\omega_i, \omega_{\neg i}) = J\omega_i m_i \tag{2}$$

where $m_i = \frac{\sum_{j\neq i}\omega_j}{N-1}$ represents the average choice of the other players, and $J(>0)$ represents the *social factor*, which weighs social utility relative to the direct private-value payoff. To be very clear on terminology: a higher *social factor*, J, increases i's (positive) *social value* if her behavior conforms to the majority choice, or decreases her (negative) social value if her behavior is in the minority. Notice that the inclusion of the social value in the payoff ensures that an individual is automatically punished if she deviates from the behavior of others. This accords with the sanctions from deviations of Bicchieri's (2017) definition of social norms.

In the model, total social value is maximized when all individuals coordinate on the same choice, and expected total private value is maximised when all individuals make the choice with the higher common value u_{ω_i}. Therefore, expected total welfare is also maximized when all individuals coordinate on this 'good' choice.

2.2 Equilibria of the stage game with full information on the common values

An equilibrium of the game can be characterized by ρ^*, the expected proportion of the group choosing $\omega_i = -1$, such that in expectation no player would be better off changing her choice. It will be useful to write this in terms of the equilibrium average choice of the group, $m^* = \frac{1}{N}\sum_{i=1}^{N}\omega_i \in [-1, 1]$. Then the equilibrium can be written as a rescaling of the support of the average choice from $[-1, 1]$ to $[0, 1]$:

$$\rho^* = \frac{1 - m^*}{2} \tag{3}$$

Players cannot *ex ante* observe m_i but instead must base their decision on their beliefs about the average group choice:

$$m_i^e = \frac{\sum_{j\neq i}\mathbb{E}_i(\omega_j)}{N-1} \tag{4}$$

where $\mathbb{E}_i(\omega_j)$ represents i's expectation over j's choice. In equilibrium, players' expectations are consistent with how others play the game. It is convenient to define $d = u_{(-1)} - u_1$ as the difference in common values and $d_i = d + \epsilon_i(-1) - \epsilon_i(1)$ as the difference in private values for player i. For example, d_i represents i's net private preference for not smoking in the absence of peer effects, while d represents the average private preference for not smoking in the group.

We are only interested in situations in which social interactions affect behavior (in expectation), and so we restrict our analysis to the region $-2J \leq d \leq 2J$.

Proposition 1 *An equilibrium is characterized by the common threshold decision rule "Choose $\omega_i = -1$ if and only if $d_i > c^*$", where the common threshold is $c^* = 2Jm^*$. An equilibrium expected average choice level of the group, m^*, solves:*

$$m^* = 2F(2Jm^* - d) - 1 \tag{5}$$

where F is the CDF of the difference in private shocks.

We relegate the proof of Proposition 1 to the "Appendix".

Equation (5) is the stage-game equilibria condition for the expected average choice level, corresponding to a common threshold c^*, for any given distribution of shocks. This is a minor generalization of Brock and Durlauf (2001).[7] The threshold c^* depends both on a player's beliefs about group behavior as well as the (fixed) social value strength. It follows that a player i maximizing her expected utility chooses $\omega_i = -1$ if $d_i > 2Jm_i^e$.

There exists at least one equilibrium and, for strictly unimodal distributions, at most three equilibria satisfying the equilibrium condition in (5).[8] The number of equilibria depends on both the social factor and the difference in common values: multiple equilibria exist only when J is sufficiently large relative to d. In such cases, and adopting for convenience the notation of (3), two stable equilibria close to the poles $\rho_-^* \approx 0$ and $\rho_+^* \approx 1$ emerge.[9]

When $d > 0$, we call the equilibrium ρ_+^* the 'good' norm, which, as previously shown, maximizes expected total welfare. When $d < 0$, ρ_-^* is the good norm. In each case, the other stable equilibrium is the bad social norm, when it exists. A bad norm is only present when multiple equilibria exist. It is noteworthy that it is not required that all or even any of the players have a private value preference for a particular choice for it to exist as a pure equilibrium.

2.3 The stage game with incomplete information on the common values

Players in the stage game with full information on the common values know both the distribution generating the private shocks for all individuals and the common values for each choice. In practical applications, people may not have such detailed information. In the game with incomplete information on the common values, players do not separately observe the common values or their individual shocks, but rather the combined private value $v_i(\omega_i) = u_{\omega_i} + \epsilon_i(\omega_i)$. Given that players receive no information at all about how the common values are determined, it is impossible to derive the theoretical benchmarks of the approach of Brock and Durlauf (2001) without making further assumptions.

[7] In Brock and Durlauf (2001) the authors assume that shocks follow an extreme value distribution. The convenient properties of this distribution allow for analytical computation of rational expectations equilibria from the symmetry of N expectations equations.

[8] Proofs are discussed in detail in (among others) Brock and Durlauf (2001) and Rothenhäusler et al. (2015).

[9] Recall that ρ^* is the expected proportion of the group choosing $\omega_i = -1$. Due to the continuous distribution of the private shocks across all possible values on the real axis, there is always a positive probability of a private difference $|d_i| > 2J$, and so the expected equilibrium proportions are never exactly at the poles 0 and 1. With some abuse of terminology, a 'mixed-proportions' equilibrium $\rho_=^* \in (\rho_-^*, \rho_+^*)$ also exists. In a setting where the parameter space is such that three equilibria exist, the equilibria at the poles are stable whereas the mixed-proportioned equilibrium is unstable. Small perturbations in players' expectations will move players away from this equilibrium.

Notice that the game with full information on the common values does not give the phenomenon of pluralistic ignorance a good shot: because players know the common values, they can be quite certain about which choice will be preferred by the majority. Pluralistic ignorance receives a fairer shot in the game with incomplete information on the common values, which may in particular be relevant to situations in which there is uncertainty about the preferences of others.

2.4 Do information and communication affect the likelihood of bad norms?

The analysis of the full information stage game indicates that both 'good' and 'bad' norms can exist as equilibria so long as the scale of social payoffs is sufficiently large with respect to the direct incentives. It does not provide any guidance on predicting which equilibrium will be selected. For the game that we study in this paper, risk dominance selects the 'good' equilibrium. The opportunity cost for player i deviating from the good equilibrium is $2J + d_i$, while the opportunity cost for player i deviating from the bad equilibrium is $2J - d_i$, and therefore the good equilibrium risk-dominates the bad equilibrium.[10] Our intuition deviates from the prediction of risk dominance here. Our conjecture is that people will continue to play according to an equilibrium after it has turned into a bad equilibrium if the social factor J is sufficiently large. This intuition is based on previous experimental work on belief learning that shows that subjects' beliefs about how other subjects will behave are clearly correlated with how these subjects behaved in the past (e.g., Cheung and Friedman 1997; Offerman et al. 2001).

The Brock and Durlauf (2001) approach cannot shed light on the question of whether full information on the common values is an important factor for the emergence and persistence of bad norms. On the other hand, the psychology literature around pluralistic ignorance argues that partial ignorance of the common values is a necessary condition for the phenomenon to occur (Prentice 2007; Bicchieri 2005; Sherif 1936). Recent evidence has found that pluralistic ignorance may play a role in the persistence of gender norms, and that correcting misperceived beliefs about group preferences can be an effective intervention (Bursztyn et al.2018). Similar interventions in other disciplines have shown promising effects on issues of college binge drinking (Schroeder and Prentice 1998), tax compliance (Wenzel 2005) and HIV prevention (Chernoff and Davison 2005). These results suggest that the information provided to the players in our experiment is an interesting treatment variable.

Another interesting factor that is motivated by past experiments but is not captured by the model is the role of communication. Communication may play a dual role in our game. It may not only help players share information about which choice they prefer but it may also help players to coordinate expectations on the same equilibrium. From this perspective, communication may have an even more positive

[10] This result depends on the linear payoff function used in our and Brock and Durlauf's (2001) model. To derive the condition for risk dominance in our game, we used the procedure described in Sect. 3.1 of Keser et al. (2012), who apply Harsanyi and Selten's (1988) tracing procedure to a technology-adoption game.

Table 1 Treatments

Treatment	Social factor J	Group size N	#Groups	Bulletin board?	Common values?
SmallWeak	4	6	8	No	No
SmallStrong	8	6	8	No	No
BigWeak	4	11	7	No	No
BigStrong	8	11	7	No	No
Communication	8	6	8	Yes	No
Full information	8	6	8	No	Yes

effect than full information. In our experiment, we are particularly interested in anonymous signalling that one might expect from posting on internet bulletin boards or social media. Online platforms can facilitate cost-free and anonymous communication to a wide audience, allowing individuals with a private interest in changing the status quo to signal their support for change in a broad manner without fear of punishment.[11] While this cheap talk is non-binding, it can also be thought of as shifting the focus away from historical precedent and towards illuminating present group preferences.[12]

The theory provides a framework that allows us to study conditions under which bad norms can emerge and persist. To shed further light on this, we turn to the lab.

3 Experimental design

The computerized experiment was run at the CREED laboratory of the University of Amsterdam. Subjects read the instructions at their own pace and then had to successfully answer some control questions before they could proceed. In the experiment, subjects earned points that were converted at the end of each session at an exchange rate of five points for 1 euro cent (500 points = 1 euro). At the start of the experiment, each subject was randomly assigned to a group and participated in 50 rounds of the game. Subjects were not told how many rounds the game would last. Points were summed over the 50 rounds and the final game earnings were paid privately. In addition, subjects received a show-up fee of 3 euros.

[11] Recent examples that have been studied include the role of social media and the internet in the 2011 'Arab Spring' uprisings (Lim 2012) and in promoting women's empowerment in India (Loiseau and Nowacka 2015).

[12] Andreoni et al. (2017) find a positive effect of communication on equilibrium selection in a similar environment. Choi and Lee (2014) find that coordination is enhanced by allowing communication in networks. However, in their experiment the roles of implicit agreement and punishment from deviations are necessary for improving coordination. Ochs (2008) shows that the effect of communication can differ in different coordination games; interestingly, this paper also highlights the role of past precedent, a mechanism that in our experiment corresponds to the strength of the bad norm.

Recruitment was conducted at the University of Amsterdam. Each subject participated in only one session of the experiment. Each session took approximately one hour. Multiple groups were run in each session, but the composition of the groups themselves remained constant. In total, 346 subjects participated in 19 sessions, and earned on average 14.30 euros (SD 2.00), including the show-up fee.

There were six treatments in total (Table 1). We start with a description of the incomplete information treatments. The game used in the experiment featured 50 rounds of the stage game of the model described in the previous section, but presented in a more subject-friendly manner. In each round players made an individual choice between two 'doors', *A* and *B*, from which they could earn points. An individual's payoff depended both on her *private value* and her *social value*. Each door's private value, which an individual observed before making the choice, consisted of the sum of that door's common value and an individual shock. Group members could not observe the components of their private values, but they knew both that the common values were the same for all group members in a given round, and that all shocks were randomly drawn from a standard normal distribution.

Social value was determined by the proportion of other group members who made the same choice as an individual, scaled by a social factor; if an individual was in the minority, the social value was negative. Specifically, the social value to a participant was formulated in terms of the number of points she would gain (lose) for each group member who made the same (different) choice as her in a given round.

After the choices by all subjects were submitted in a given round, the payoffs were presented along with information about the number of other group members who chose each door. The experiment then continued to the next round, and subjects saw their new private values for the doors.

The common door values used in the experiments were randomly generated, subject to certain criteria. Specifically, unknown to the subjects,

- Door A was initially preferred by a large margin (roughly 6 points).
- Common values of each door could change by a maximum of 1 point in each new round.
- Door A remained preferable until round 25, after which Door B overtook Door A.
- From round 40 until the end of the session, Door B held a positive difference over Door A of approximately 2 points.

These stipulations were designed to create an environment in which in the first half of the session, a social norm of choosing Door A could emerge, which would then consistently be the socially inefficient choice in the second half. Figure 1 shows how the common door values developed over time in each group of each treatment.

The group sizes ($N = 6, 11$) were chosen to make it easier for subjects to calculate the potential social values, which required considering fractions of 5 or 10. The social factors ($J = 4, 8$) were chosen so as to predict opposite equilibria in computer simulations in which agents assigned equal weights to both the existing norm and their own private information in forming expectations. The actual presentation of the instructions multiplied $u_{\omega_{it}}$, J and $\epsilon_{it}(\omega_{it})$ from the theoretical model by 10 so that

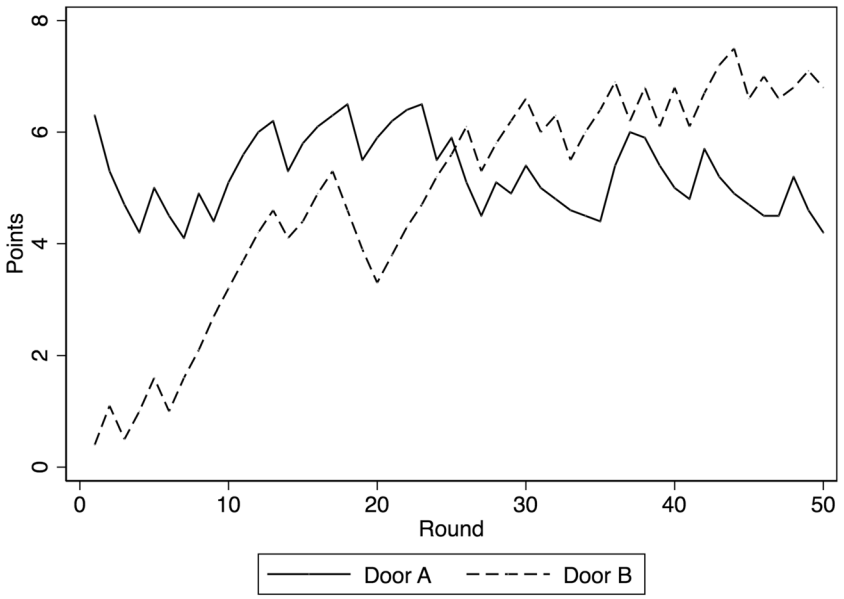

Fig. 1 Common door values. *Notes*: For participants in the laboratory experiment, all values were multiplied by 10

subjects did not have to calculate decimals. We continue to use the unmultiplied values in the rest of the paper for consistency.

To make things easier for subjects to understand, the linear nature of the social value was explained in terms of the number of points earned per other player making the same choice. Wording was of the form: "You gain X points for every person who makes the same choice as you, but you lose X points for every person who makes the opposite choice to you", where we substituted the appropriate value for X depending on the treatment. Private shocks were randomly drawn from $\sim \mathcal{N}(0, 1)$ for each individual, door and round. Realizations of private shock distributions for each individual were matched for treatments with the same group size. That is, each of the 8 groups in *SmallWeak* had a matched group in *SmallStrong*, *FullInformation* and *Communication* with the same private shocks distributed across group members, doors and rounds, and likewise for the 7 groups in each of the larger treatments.

All treatments made use of the experimental variant of the game described above. In the *Full Information* treatment we replicated the parameters of the *SmallStrong* treatment ($N = 6, J = 8$), but gave subjects full information about the true distribution of others' private preferences. Specifically, subjects could precisely see the decomposition of their private values into the common values and their own personal shocks for each door in every round. Subjects were not informed of the specific shocks for the other group members, but knew the distribution generating the draws.

The *Communication* treatment replicated the information structure and parameters of the *SmallStrong* treatment, but allowed subjects to communicate. In every

round before they chose their door, each subject could express her intention on a 'Bulletin Board'. Posts on the Bulletin Board were anonymous. Subjects were informed that there was no obligation to honor a post, and that it was also possible not to post anything. After everyone had made their decisions about posting for that round, group-members saw the total number of posts (or 'intentions to choose') for Door A and Door B before they actually made their final choice of door.

Table 1 summarizes the main features of all treatments. These were varied between subjects, with the three additional treatments to those listed above being based on combinations of the two parameters of interest: the social factor and the group size. In each round of each treatment, subjects' screens displayed the round number, the cumulative earnings, the private values for each door, a choice button for Door A or Door B to be submitted, and a history footer. The history footer contained the total history of the proportion of other group members making each choice for every completed round. At the end of round 50, subjects filled out a short questionnaire before they were paid.

The experimental design mimics the theoretical model in that an individual i receives a payoff in round t of:

$$V_{it}(\omega_{it}) = u_{\omega_{it}} + J\omega_{it}\sum_{j\neq i}\frac{\omega_{jt}}{(N-1)} + \epsilon_{it}(\omega_{it}), \quad \omega_{it} \in \{-1, 1\} \tag{6}$$

Here $\omega_{it} = 1$ is defined as individual i choosing Door A and $\omega_{it} = -1$ as choosing Door B. $u_{\omega_{it}}$ is the common value from the chosen door in round t, which is the same for every individual who chooses $\omega_{jt} = \omega_{it}$ and whose realizations for the whole experiment were generated as previously discussed (displayed in Fig. 1). J is the social factor (see Table 1), and $\epsilon_{it}(\omega_{it})$ is i's door-specific shock in round t (i.i.d drawn from $\sim \mathcal{N}(0, 1)$).[13]

3.1 Static equilibria and hypotheses for the experimental parameters

Door A was the group welfare-maximizing or 'good' norm in rounds 1–25 and became the 'bad' norm in rounds 26–50, while for Door B the situation was reversed. Given that we only have theoretical benchmarks for the full information case, we let these guide our experimental parameter choices of all treatments. The parametrization of the experiment ensured that both pure Bayesian Nash equilibria satisfying the threshold condition of Proposition 1 were supported in rounds 25–50, when Door A became a bad norm, in each treatment (Fig. 2). The equilibria ρ^*_+ and ρ^*_- were extremely close to 1 and 0 in every case, and a mixed equilibrium $\rho^*_= \in (0, 1)$ was also always present. Notice that for each round, the difference in group payoffs between the good equilibrium ρ^*_+ and the bad equilibrium ρ^*_- are the same for $J = 4$ and $J = 8$.[14]

[13] As previously mentioned, actual payoffs were multiplied by 10 when presented to subjects. Instructions and an example screenshot are displayed in the "Appendix".

[14] The parametrization for $J = 4$ meant that Door A was the sole equilibrium in round 1.

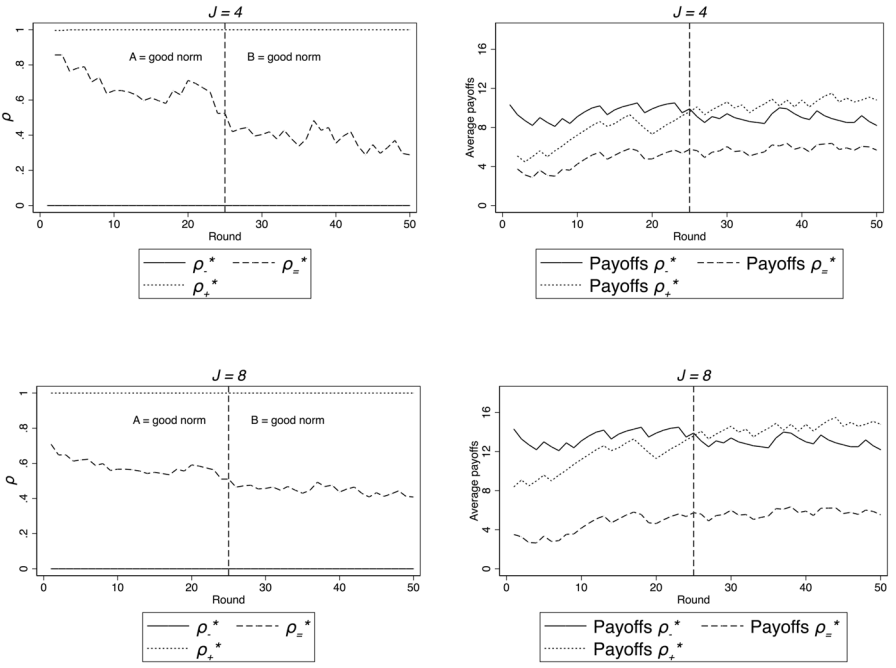

Fig. 2 Experiment equilibria and payoffs. *Notes*: The left panels depict the solutions in each round to the equilibrium condition for the full information stage game (Proposition 1) for the parameters used in the experiment, where ρ is the proportion of the group choosing Door B. A maximum of three solutions exist: two equilibria near the boundaries 0 and 1, which we label ρ^*_- and ρ^*_+, and an unstable mixed equilibrium $\rho^*_=$. In each round, $\rho^*_- < .01$ and $\rho^*_+ > .99$. The right panels depict the corresponding average individual payoffs from each equilibrium. Door A was the welfare-maximizing group choice in rounds 1–25 and this switched to Door B in rounds 26–50, as shown by the common values in Fig. 1. For round 1 of groups with $J = 4$, only one equilibrium (ρ^*_+) existed

The full information stage-game analysis reveals that both pure equilibria exist given our parameters, but it is silent about which will be selected. However, given a norm of Door A emerging by round 25 and some distribution of expectations, our intuition was that the now bad norm is more likely to persist by round 50 when the social factor J is larger, because a larger J makes it more costly for subjects to experiment to see if others are also willing to deviate from the norm. While the equilibrium condition does not depend on group size, an equilibrium is more robust to perturbations when N is larger in the sense that realized shocks are less likely to breach a 'tipping proportion' (see "Appendix"). This equilibrium feature supported our conjecture that the bad norm is sooner escaped in smaller groups. Notice, however, that evidence from psychology experiments on conformity is mixed (e.g., Asch 1952; Mann 1977; Wilder 1977; Bond 2005).[15] Conditional on J and N, the model

[15] Economics experiments involving the minimum-effort game have found a strong negative effect of group size on coordination, but this game is fundamentally different to our game in this respect, as discussed in Sect. 1. In the minimum-effort game, subjects are punished if one group member chooses a lower effort level, whereas in our game, punishment (a lower social value) depends on the proportion of others making the opposite choice. See also Weber (2006).

does not differentiate between the different setups of *FullInformation*, *SmallStrong* and *Communication* in its predictions.

On the other hand, past empirical evidence motivate two further predictions. The psychological literature on pluralistic ignorance argues that uncertainty over the true distribution of private preferences is an important condition for this phenomenon to exist (Prentice 2007). Therefore, we expected the bad norm to be less likely to persist when subjects received complete information about the common values and shocks *FullInformation* than in the corresponding incomplete-information treatment *SmallStrong*. We also expected groups in the *Communication* treatment to be similarly more successful in escaping from the bad norm because communication allowed subjects to coordinate their expectations of which choice would attract the majority in the group (Andreoni et al.2017; Choi and Lee 2014; Ochs 2008).

Our experimental outcome of interest is whether groups can eventually escape a bad norm. More precisely, the main outcome variable is ρ_{50}, the proportion of the group choosing Door B in the final round of the game. Below we summarize the main hypotheses that our treatments allow us to test.

Hypothesis 1 Role of information

$$\rho_{50} \textit{ in } \textit{FullInformation} = \rho_{50} \textit{ in } \textit{SmallStrong}$$

Hypothesis 2 Role of communication

$$\rho_{50} \textit{ in } \textit{Communication} = \rho_{50} \textit{ in } \textit{SmallStrong}$$

Hypothesis 3 Role of social factor

$$\rho_{50} \textit{ in } \textit{SmallWeak} = \rho_{50} \textit{ in } \textit{SmallStrong} \quad \textit{and}$$
$$\rho_{50} \textit{ in } \textit{BigWeak} = \rho_{50} \textit{ in } \textit{BigStrong}$$

Hypothesis 4 Role of group size

$$\rho_{50} \textit{ in } \textit{BigStrong} = \rho_{50} \textit{ in } \textit{SmallStrong and}$$
$$\rho_{50} \textit{ in } \textit{BigWeak} = \rho_{50} \textit{ in } \textit{SmallWeak}.$$

4 Results

We present the results in two parts. Section 4.1 investigates the roles that information, communication and pluralistic ignorance play for the persistence of bad norms. Section 4.2 provides the results of the other treatments, focusing on the parameters of the Brock and Durlauf (2001) model. It sheds light on what factors are important when there is incomplete information.

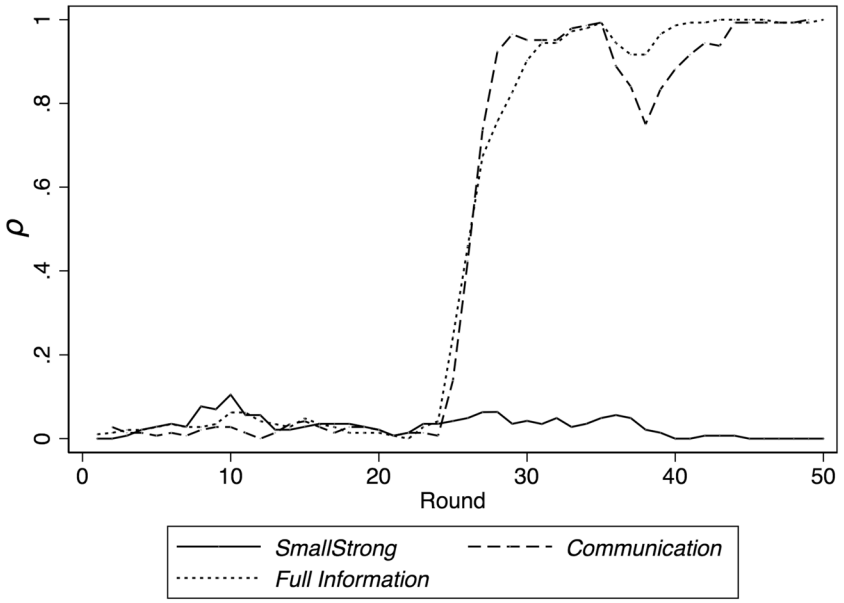

Fig. 3 Average round-by-round group choice for $N = 6, J = 8$, including anonymous communication and decomposed private values (full information) treatments. *Notes*: Each treatment line depicts the average group proportion choosing Door B across all groups in the treatment. After round 25 , coordination on Door A represents a bad social norm. Lines have been smoothed via a three-round equally weighted moving average

4.1 Pluralistic ignorance: the role of information and communication

A comparison of the *SmallStrong* treatment, the *Full Information* treatment and the *Communication* treatment allows us to shed light on pluralistic ignorance. Figure 3 shows when groups on average switch to the good norm in these treatments (if they do). Notably, the bad norm remained until round 50 in every group in the *SmallStrong* treatment, where the strong social factor provides favorable conditions for the persistence of bad norms. Providing full information on the decomposition of common values and private shocks in the game with the same parameters dramatically changes the picture. All groups escaped the bad norm in both the *Full Information* and *Communication* treatments, a significant improvement over groups in *SmallStrong* ($p < .01$ for both pairwise rank-sum tests; see Table 2). In the *Full Information* treatment, most groups switched to the good norm in round 26, immediately after the common values shifted towards Door B, and all groups switched to the good norm by round 28. Remarkably, all groups also escaped the bad norm almost immediately in the *Communication* treatment, despite the shift of the common values being unobserved. These results accord with psychological theories of social norms that propose that payoff uncertainty of other group members is a crucial ingredient for bad norm persistence.

While we did not explicitly measure expectations, we can identify extreme circumstances where we might observe behavior consistent with pluralistic

Table 2 Key performance indicators by treatment

	Treatments	ρ_{50}	$\bar{\rho}_{(45-50)}$	$\bar{\rho}_{(t\geq 26)}$	$\bar{\rho}_{all}$	$\bar{t}_{switch}$
	Full Information	1.00	1.00	.93	.48	26
	Communication	1.00	1.00	.92	.47	27
	SmallStrong	.00	.00	.03	.03	–
	SmallWeak	.65	.62	.46	.26	31
	BigStrong	.03	.02	.02	.02	–
	BigWeak	.47	.36	.26	.14	39
Testing information uncertainty	FI versus SS	.00***	.00***	.00***	.00***	–
Testing communication	C versus SS	.00***	.00***	.00***	.00***	–
Testing social factor	SW versus SS	.00***	.00***	.01***	.01***	–
		(.01***)	(.01***)	(.02**)	(.02**)	(–)
	BW versus BS	.02**	.04**	.11	.06*	–
		(.03**)	(.08*)	(.22)	(.13)	(–)
Testing group size	SW versus BW	.46	.41	.30	.30	.17
		(.92)	(.83)	(.61)	(.61)	.17
	SS versus BS	.12	.02**	.82	.56	–
		(.23)	(.03**)	(1.00)	(1.00)	(–)

In the upper panel, values are averages of the group values within each treatment. ρ_{50} is the final group proportion choosing Door B. $\bar{\rho}_{(45-50)}$ is the average ρ across the last six rounds. $\bar{\rho}_{all}$ is the average ρ across all rounds. $\bar{\rho}_{(t\geq 26)}$ is the average ρ from round 26, when the common value of Door B becomes larger than that of Door A. $\bar{t}_{switch}$ is the average switching time, considering only those groups that switched to Door B by round 50. In the lower panels, *p*-values are derived from Mann–Whitney rank sum tests. Bonferroni-adjusted *p*-values for the multiple comparisons in testing the social factor and group size are given in parentheses. In the tests, each group yields one observation. Full indicators by group are found in the "Appendix"

*** $p < 0.01$, ** $p < 0.05$

ignorance. Specifically, if all individuals in a group have a private value of Door B exceeding that of Door A in a particular round of the experiment, but *all* group members choose Door A ($\rho = 0$), the group is said to exhibit *total pluralistic ignorance*. Such incidence represents the worst-case scenario of conformity from a social welfare perspective; in fact, if social value is ignored, any other combination of choices would be a Pareto improvement. In the experiment the number of rounds in which total pluralistic ignorance could potentially exist is naturally higher for smaller groups, as groups with more individuals are more likely to produce at least one group member realizing extreme private shocks. Figure 4 compares the number of potential rounds of total pluralistic ignorance to those that eventuated in the experiment. This reveals a strong social factor effect. *SmallStrong* and *BigStrong* saw total pluralistic ignorance in, respectively, an average of 87% and 81% of each treatment's potential rounds, while for *SmallWeak* and *BigWeak* the average frequencies were 27% and 31%. On the other hand, in both the *Communication* and *FullInformation* treatments, no group ever exhibited total

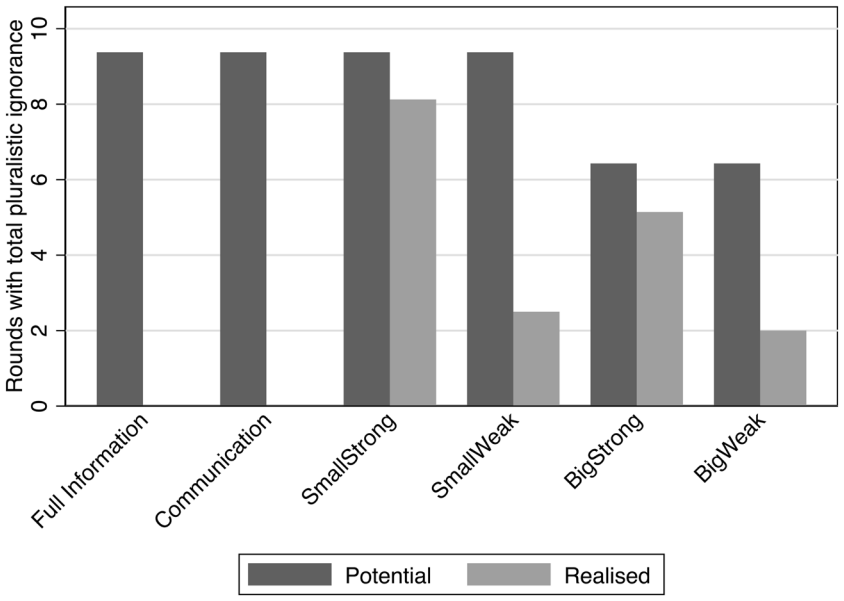

Fig. 4 Mean potential and realized rounds of total pluralistic ignorance. *Notes*: A 'total pluralistic ignorance' round is defined as a round t in which all players receive $d_{it} > 0$ and subsequently choose Door A ($\rho_t = 0$). Amounts are averages per group out of a total of 50 rounds

pluralistic ignorance for any round, a significant difference to the behavior in the comparison treatment *SmallStrong* ($p = .00$ for both pairwise rank-sum tests).

In the *Communication* treatment, only two of the 48 participants chose not to use the Bulletin Board at all; of the rest, most subjects took the opportunity to post in every round. Moreover, the collection of posts on the Bulletin Board was overwhelmingly indicated as the primary means of expectation formation in the answers to the questionnaire. Figure 5 presents the average number of announcements to opt for Door B together with the actual choices for Door B as the rounds unfolded. For all eight groups, the switch in average group indications from Door A to Door B coincided with the shift in the difference in common values. Interestingly, all participants exploited the anonymity by acting contrary to their posted indication in at least one round (mean = 5.3 rounds, SD = 2.4).

The above analysis allows us to reject our first two hypotheses regarding the effects of full information on preferences and communication:

Result 1 ρ_{50} *in FullInformation* $> \rho_{50}$ *in SmallStrong*

In agreement with pluralistic ignorance, bad norms are more likely to persist when subjects are uncertain about others' preferences than when subjects are fully informed.

Result 2 ρ_{50} *in Communication* $> \rho_{50}$ *in SmallStrong*

Bad norms are less likely to persist when subjects can communicate.

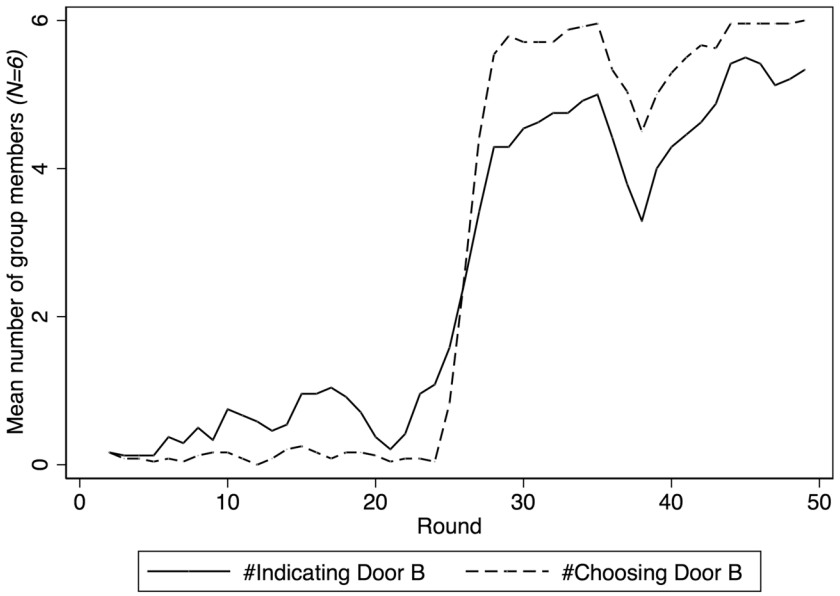

Fig. 5 Indications and actual choices in the *Communication* treatment, by group. *Notes*: Average round-by-round group indications for Door B versus actual choices in the *Communication* treatment. Treatment parameters were: $N = 6$, $J = 8$. Almost all subjects in a group posted their intentions in every round (mean = 5.6 group members, SD = 0.6). Lines have been smoothed via a three-round equally weighted moving average

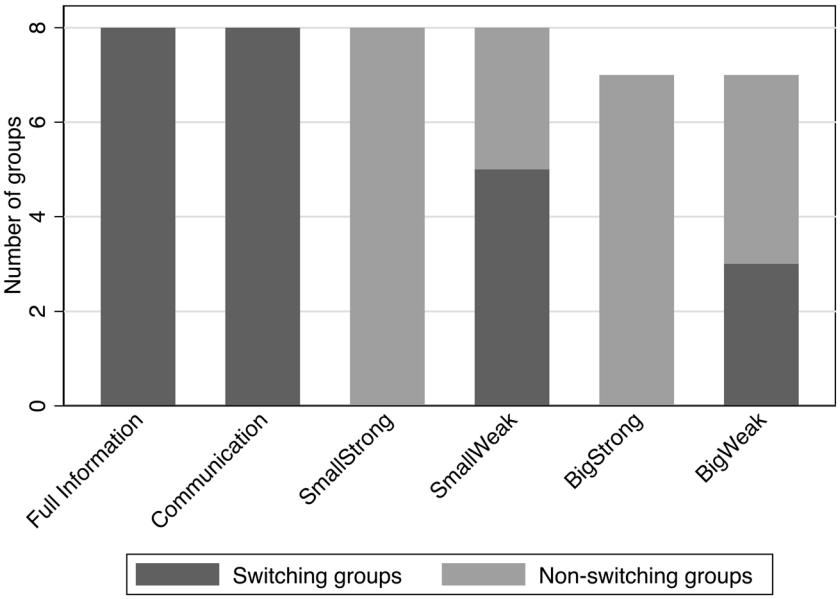

Fig. 6 Switching groups by treatment. *Notes*: 'Switching' is defined as more than half of the group choosing Door B in round 50 ($\rho_{50} > 0.5$)

4.2 The roles of the social factor and group size when there is incomplete information

Figure 6 displays the frequency of groups escaping the bad norm in the treatments with incomplete information. None of the groups with the strong social factor ($J = 8$) switched to Door B by round 50, regardless of group size. When the social factor was weakened to $J = 4$, five out of the eight groups (62.5%) in *SmallWeak* switched to Door B, while three out of seven (42.9%) did the same in the *BigWeak* treatment. The simulations of the theoretical model for the common values, shocks and treatments used in the experiment also predicted a slight favoritism for *SmallWeak* compared to *BigWeak* for the sequence of common values used.

Table 2 demonstrates that the descriptive statistics and tests of the data partitioned by treatment are similar when we define our outcome by different measures, such as the average ρ across all rounds, the final rounds, or rounds 26–50 (the rounds after which the common value of Door B overtakes that of Door A). Detailed proportions for the 46 individual groups can be found in the "Appendix". For each individual group, the average group choice stuck closely to the two full information stage-game equilibria of $\rho = 0$ and 1 across the rounds; groups spent few rounds in the socially destructive mixed proportions around $\rho = 0.5$. For the groups that eventually escaped the bad norm, once approximately a third of the group had simultaneously chosen Door B the group generally took little time in reaching the more favorable equilibrium.

The third key result reflects the strength of the social factor. We reject the hypothesis that bad norms are unaffected by the social factor under incomplete information and find that they are more likely to persist when the social factor is larger. The upper panel of Table 2 clarifies that the social factor has a substantial impact on the proportion switching to the good door in the latter part of the experiment. When J is strong, the group choice of Door A persisted after it had become the bad choice. The lower panels of Table 2 show the extent to which the results differ systematically across treatments. A weak social factor significantly increases various measures of ρ for both $N = 6$ and $N = 11$. The result is further illustrated in Fig. 7. Only groups with the weaker social factor switched their overall door preference after round 25.

Result 3

ρ_{50} *in SmallWeak* $> \rho_{50}$ *in SmallStrong and*
ρ_{50} *in BigWeak* $> \rho_{50}$ in *BigStrong*

With incomplete information, bad norms are more likely to persist when the group's social factor is strong.

The final key result concerns the role of group size. Here, we cannot reject the hypothesis that group size has no effect on long-run persistence. The tests on the long-run (round 50) effects of group size reported in the lower panel of Table 2 are generally insignificant. When only the weaker social factor groups are considered, smaller groups were more successful at escaping the bad norm across all of the outcome measures in Table 2, but these differences are not statistically significant.

Result 4

ρ_{50} *in BigStrong* $= \rho_{50}$ *in SmallStrong and*
ρ_{50} *in BigWeak* $= \rho_{50}$ *in SmallWeak*

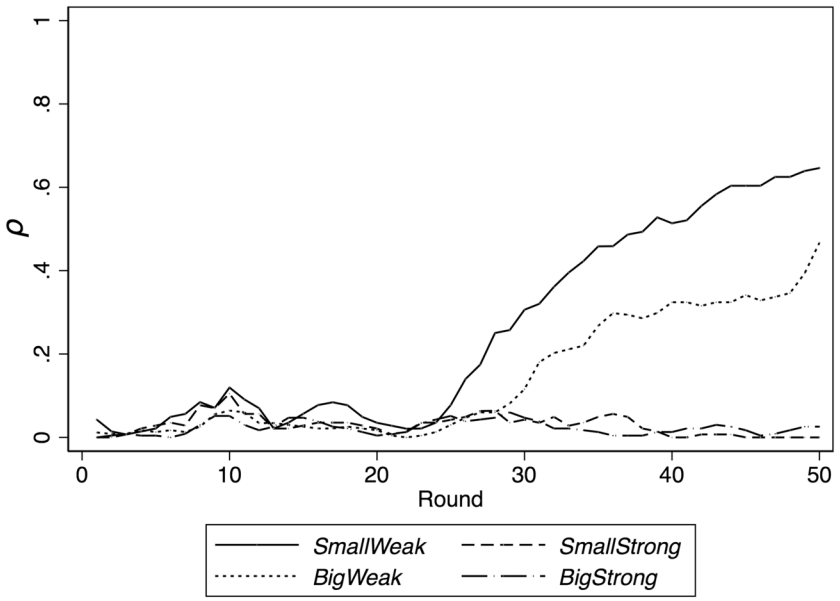

Fig. 7 Effect of the social factor and group size on group choice. *Notes*: The figure shows the average round-by-round group choice for treatments with information uncertainty and no communication. This highlights the effect of the social factor J and group size N on group choices. Each treatment line depicts the average group proportion choosing Door B across all groups in the treatment. Lines have been smoothed via a three-round equally weighted moving average

With uncertainty, the persistence of bad norms does not depend on group size in the long run.

Nevertheless, the graphical representation of round-by-round pooled data presented in Fig. 7 suggests that groups of size $N = 6$ that switched to Door B generally did so earlier than the switching groups of size $N = 11$. This difference is not statistically significant, though this is possibly due to the small sample size of switching groups ($\bar{t}^{SW}_{switch} = 31.4, \bar{t}^{BW}_{switch} = 39.0, t(6) = -1.59$, one-sided $p = .09$).

5 Discussion

When there is uncertainty about the preferences of others, bad norms can persist in the laboratory. Bad norms emerge in our experiment as a result of a good equilibrium gradually becoming a bad equilibrium in a coordination game due to changing payoffs over time. Once established, these bad norms can persist so long as the personal incentives to deviate are small and the social factor is strong.

The most important insight from our experiment is that a strong interdependence of payoffs is a necessary but not sufficient condition for the persistence of bad norms. That is, when a strong social factor is paired with full information about the preferences of others, bad norms disappear. This is consistent with insights from psychology, specifically that uncertainty about the true distribution of the private

preferences of group members is a necessary condition for pluralistic ignorance to persist. We reason from our empirical findings that an important condition for bad norm persistence is uncertainty about others' private preferences.

When there is incomplete information, we find that costless communication can weaken the persistence of bad norms. This is consistent with a broad experimental literature on the cooperative benefits of communication (Ostrom 2000). The successful effect of the bulletin board treatment in our experiment may suggest investigation of similar social norm interventions in the field, such as the use of social media or other online tools.[16] Another interesting result from the experiment is that smaller groups that escaped the bad norm did so faster than larger groups, although the prospects between differently sized groups converged by the end of the game.

The theory of Brock and Durlauf (2001) does not yet capture the dynamics of norm persistence in the presence of pluralistic ignorance, and in particular the mechanics by which a group can move towards one equilibrium or another. A dynamic model would be useful both to predict which equilibrium is likely to persist in the long run, and for cases where the welfare effects along the equilibrium path are especially policy-relevant. Broadly speaking, there are two classes of models to consider, depending on whether agents are backward- or forward-looking.[17] An example of the former approach is Lim and Neary (2016), who find evidence in the language game that individuals behave consistently with a myopic best-response learning rule based on the previous period's outcomes. The second approach may be better suited to incorporate the implications of pluralistic ignorance, such as by describing a dynamic belief formation function that depends on the uncertainty of information. When there is full information, individuals place less weight on the previous norm and more on the current preferences of others when forming their expectations about future group behavior. Correspondingly, bad norm persistence requires uncertainty about others' preferences, in agreement with pluralistic ignorance.

Our paper suggests other interesting avenues for future research. Our experimental design automatically monetizes all payoffs that derive from the behavior of the self and others. Further research into applications that feature internalized social payoffs could consider directly triggering group identity in the laboratory, along the lines of Chen and Chen (2011), Charness et al. (2007) etc . What a more natural setting of this nature loses in robustness would be compensated by adding support to the behavioral foundations of the modeling of bad social norms proposed in this paper. Our results also motivate a need for further tests in the field, and suggest that bad norm interventions that target reducing uncertainty are worthy of consideration.

Acknowledgements We thank the editor, two referees, seminar attendees from the University of Nottingham, the University of East Anglia and the University of Amsterdam (Grant No. 201212170412), as well as Cars Hommes, Adriaan Soetevent, Swapnil Singh and Sabina Albrecht for helpful comments.

[16] For instance, recent experimental evidence suggests that using Twitter as an intervention tool can be effective in combating norms of racial harassment (Munger 2017).

[17] We thank an anonymous referee for these suggestions.

Appendix

Table of results

See Table 3.

Table 3 Key performance indicators by group

Group	Treatment	N	J	ρ_{50}	$\bar{\rho}_{(45-50)}$	$\bar{\rho}_{all}$	$\bar{\rho}_{(t\geq 26)}$	t_{switch}	Earnings (€)
1	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.49**	**.93**	**26**	**12.32**
2	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.47**	**.91**	**26**	**12.38**
3	**Full Information**	**6**	**8**	**1.00**	**.97**	**.49**	**.94**	**26**	**12.49**
4	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.51**	**.99**	**26**	**13.49**
5	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.49**	**.96**	**26**	**13.22**
6	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.44**	**.87**	**27**	**12.48**
7	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.46**	**.88**	**28**	**12.04**
8	**Full Information**	**6**	**8**	**1.00**	**1.00**	**.50**	**.98**	**26**	**13.45**
9	**Communication**	**6**	**8**	**1.00**	**1.00**	**.44**	**.87**	**27**	**13.42**
10	**Communication**	**6**	**8**	**1.00**	**1.00**	.49	**.95**	**27**	**13.65**
11	**Communication**	**6**	**8**	**1.00**	.97	**.50**	**.93**	**26**	**11.98**
12	**Communication**	**6**	**8**	**1.00**	**1.00**	**.49**	**.98**	**26**	**13.70**
13	**Communication**	**6**	**8**	**1.00**	**1.00**	**.44**	**.88**	**28**	**13.67**
14	**Communication**	**6**	**8**	**1.00**	**1.00**	**.44**	**.86**	**26**	**12.83**
15	**Communication**	**6**	**8**	**1.00**	**1.00**	**.45**	**.87**	**27**	**13.10**
16	**Communication**	**6**	**8**	**1.00**	**1.00**	**.50**	**.99**	**26**	**13.79**
17	SmallStrong	6	8	.00	.00	.03	.02	–	12.39
18	SmallStrong	6	8	.00	.00	.05	.07	–	11.77
19	SmallStrong	6	8	.00	.00	.02	.01	–	12.78
20	SmallStrong	6	8	.00	.00	.05	.06	–	11.77
21	SmallStrong	6	8	.00	.00	.01	.00	–	12.90
22	SmallStrong	6	8	.00	.00	.02	.00	–	12.49
23	SmallStrong	6	8	.00	.00	.03	.03	–	12.38
24	SmallStrong	6	8	.00	.00	.02	.01	–	12.72
25	SmallWeak	6	4	.00	.00	.04	.03	–	8.76
26	SmallWeak	6	4	.17	.06	.03	.03	–	8.76
27	SmallWeak	6	4	.00	.00	.03	.04	–	8.84
28	**SmallWeak**	**6**	**4**	**1.00**	**0.97**	**.27**	**.49**	**38**	**8.39**
29	**SmallWeak**	**6**	**4**	**1.00**	**1.00**	**.42**	**.79**	**30**	**8.91**
30	**SmallWeak**	**6**	**4**	**1.00**	**1.00**	**.46**	**.85**	**28**	**8.69**
31	**SmallWeak**	**6**	**4**	**1.00**	**1.00**	**.32**	**.62**	**35**	**9.21**
32	**SmallWeak**	**6**	**4**	**1.00**	**.94**	**.48**	**.84**	**26**	**8.09**
33	BigStrong	11	8	.00	.02	.02	.02	–	12.82
34	BigStrong	11	8	.00	.00	.03	.04	–	12.28
35	BigStrong	11	8	.09	.06	.02	.02	–	12.80
36	BigStrong	11	8	.00	.02	.03	.03	–	12.38

Table 3 (continued)

Group	Treatment	N	J	ρ_{50}	$\bar{\rho}_{(45-50)}$	$\bar{\rho}_{all}$	$\bar{\rho}_{(t\geq 26)}$	t_{switch}	Earnings (€)
37	BigStrong	11	8	.00	.00	.04	.04	–	12.04
38	BigStrong	11	8	.00	.00	.01	.00	–	13.09
39	BigStrong	11	8	.09	.02	.02	.02	–	12.58
40	BigWeak	11	4	.09	.02	.01	.01	–	9.07
41	**BigWeak**	**11**	**4**	**1.00**	**.98**	**.40**	**.76**	**32**	**8.97**
42	**BigWeak**	**11**	**4**	**.91**	**.41**	**.10**	**.18**	**49**	**8.45**
43	**BigWeak**	**11**	**4**	**1.00**	**.98**	**.32**	**.62**	**36**	**8.99**
44	BigWeak	11	4	.00	.00	.02	.01	–	9.00
45	BigWeak	11	4	.09	.05	.06	.09	–	8.56
46	BigWeak	11	4	.18	.09	.07	.12	–	8.48

Values are averages over group values. Earnings do not include the €3 show-up fee. ρ_{50} = final group proportion choosing Door A. $\bar{\rho}_{(45-50)}$ = average ρ across the last six rounds. $\bar{\rho}_{all}$ = average ρ across all rounds. $\bar{\rho}_{(t\geq 26)}$ = average ρ from round 26, when the common value of Door B becomes larger than that of Door A. t_{switch} is the first round in which switching groups switched to Door B. Bold rows are those groups defined as having switched to Door B by the end of the experiment

Proofs

Stage-game equilibria

It follows from the decision rule specified in Proposition 1 that, in equilibrium, we require that players prefer $\omega_i = 1$ at least as much as $\omega_i = -1$ if $d_i < c^*$, that players prefer $\omega_i = -1$ at least as much as $\omega_i = 1$ if $d_i > c^*$ and, in particular, that a player is exactly indifferent between $\omega_i = -1$ and 1 if she draws private values with a difference equal to the threshold c^*. We use this latter property of the equilibrium to endogenously calculate the threshold.

The threshold c^* depends both on an individual's beliefs about group behavior as well as the (fixed) social factor. Solving for this threshold allows us to compute a general equilibria condition that holds for any given distribution of the private shocks. Then an individual i maximizing her expected utility chooses $\omega_i = -1$ if $d_i > 2Jm_i^e$. To endogenously solve for an equilibrium, we first rewrite m_i^e as:

$$m_i^e = \frac{1}{N-1}\sum_{k=0}^{N-1}\left(\binom{N-1}{k}p^k(1-p)^{(N-1-k)}(2k-N+1)\right) \tag{7}$$

where p is the probability of a single draw of $d_i < c^*$ so that i chooses $\omega_i = 1$. Then each term in the series is the expected value for each possible value of m_i, which can be written in the form $\frac{2k-N+1}{N-1}$ for each $k \in \{0, N-1\}$.

Letting m_i^{e*} be the equilibrium expected average choice of the others in a group, corresponding to a threshold c^*, we can rewrite $c^* = 2Jm_i^{e*}$ in (7). Then

solving for an individual i drawing exactly $d_i = c^*$ with $\text{e}V_i(-1) = \text{e}V_i(1)$ allows us to solve endogenously for the expectation $m_i^{e*} = m_j^{e*} \ \ \forall i, j$:

$$m_i^{e*} = \frac{1}{N-1} \sum_{k=0}^{N-1} \binom{N-1}{k} F(2Jm_i^{e*} - d)^k (1 - F(2Jm_i^{e*} - d))^{(N-1-k)} (2k - N + 1) \tag{8}$$

At first sight, an individual's expectations appears to depend on the size of the group, N. We perform the replacements $M = N - 1$ and $F = F(2Jm_i^{e*} - d)$ for notational convenience to rewrite (8) as:

$$m_i^{e*} = \frac{1}{M} \sum_{k=0}^{M} \binom{M}{k} F^k (1 - F)^{(M-k)} (2k - M) \tag{9}$$

It can be shown that the sum of this series is independent of group size as follows: Let k be a binomially-distributed random variable with parameters $n = M, p = F$. Then $\mathbb{E}(k) = MF$ and so the right-hand side of (9) simplifies to $2F - 1$.

Thus, (8) can be rewritten as $m_i^{e*} = 2F(2Jm_i^{e*} - d) - 1$, which notably does not depend on N. Similarly, the researcher's prediction of the expected average choice level of the whole group solves:

$$m^* = 2F(2Jm^* - d) - 1 \tag{10}$$

Effect of group size

While group size does not influence the stage-game equilibria, it may still affect the probability of a group switching from a bad equilibrium to a good equilibrium in a given round. Consider a scenario in which the bad norm $\omega_{it} = 1$ is persistent on account of relatively large J and m_{it}^e, such that in the majority of rounds $\rho_{it} = 0$. Ex-ante, the probability of an individual choosing $\omega_{it} = -1$ in a given round t is $\hat{\rho}_t$, regardless of the group size. Now consider the rounds in which $0 < \rho_{it} < 0.5$; that is, the bad norm $\omega_i = 1$ is still in effect but *at least one* group member receives a private shock difference large enough to induce choosing $\omega_{it} = -1$. This likelihood is not the same across group sizes. The probability that at least one group member chooses $\omega_{it} = -1$ increases with N, and so we would expect *a higher proportion of rounds with* $\rho_{it} \neq 0$ *in larger groups* while the bad norm persists. However, the marginal effect of a group member choosing $\omega_{it} = -1$ on the overall group proportion ρ_{it} decreases with N, and so of those rounds where $\rho_{it} \neq 0$ while the bad norm persists, we would expect that ρ_{it} *is higher on average for smaller groups*.

Now, assume there is some 'tipping proportion' $\tilde{\rho}$ that, if reached after a previous equilibrium of full conformity to the bad norm ($\rho^* \approx 0$), would result in a switch to the 'good' equilibrium $\rho^* \approx 1$ with almost certainty. The tipping proportion is greater than the predicted group proportion $\hat{\rho}_t$ so that on expectation it should not be breached in a given round. Then, after a round in which $\rho_{t-1} \approx 0$, the probability of reaching the tipping proportion in round t is the probability that at least $N\tilde{\rho}$ individuals choose $\omega_{it} = -1$. From the researcher's perspective, the number of individuals choosing $\omega_{it} = -1$ follows a binomial distribution so that $N\rho_t \sim \mathcal{B}(N, \hat{\rho}_t)$ and hence:

$$\begin{aligned}\Pr\left(\rho_t \geq \tilde{\rho}\right) &= 1 - \Pr\left(\rho_t < \tilde{\rho}\right) \\ &= 1 - \sum_{j=0}^{\lfloor N\tilde{\rho} \rfloor} \binom{N}{j} \hat{\rho}_t^j (1-\hat{\rho}_t)^{N-j}\end{aligned} \tag{11}$$

where $\lfloor N\tilde{\rho} \rfloor$ is the largest integer less than $N\tilde{\rho}$.

This function does not change monotonically with N. However, some idea can be garnered as to how the probability is affected across general size increases. The binomial distribution can be approximated by a normal distribution with mean $N\hat{\rho}_t$ and variance $N\hat{\rho}_t(1-\hat{\rho}_t)$ when $N\hat{\rho}_t > 5$. Assuming this is met, equation (11) can be approximated by:

$$\begin{aligned}\Pr\left(\rho_t \geq \tilde{\rho}\right) &= 1 - \Pr\left(\frac{N(\rho_t - \hat{\rho}_t)}{\sqrt{N\hat{\rho}_t(1-\hat{\rho}_t)}} < \frac{N(\tilde{\rho} - \hat{\rho}_t)}{\sqrt{N\hat{\rho}_t(1-\hat{\rho}_t)}}\right) \\ &\approx 1 - \boldsymbol{\Phi}\left(\sqrt{N}\frac{\tilde{\rho} - \hat{\rho}_t}{\sqrt{\hat{\rho}_t(1-\hat{\rho}_t)}}\right)\end{aligned} \tag{12}$$

which, for $\tilde{\rho} > \hat{\rho}_t$, is a decreasing function of N.

When a bad norm is in effect, smaller groups are thus *generally* more likely to breach the tipping proportion in a given round. The effect of size on persistence increases slowly and not monotonically, although comparisons can be made for sizes that are not very close together. This is due to the discrete nature of the possible proportions and hence the upper sum limit $\lfloor N\tilde{\rho} \rfloor$.

References

Abbink, K., Brandts, J., Herrmann, B., & Orzen, H. (2010). Intergroup conflict and intra-group punishment in an experimental contest game. *American Economic Review*, *100*(1), 420–447. https://doi.org/10.1257/aer.100.1.420.

Abbink, K., Gangadharan, L., Handfield, T., & Thrasher, J. (2017). Peer punishment promotes enforcement of bad social norms. *Nature Communications*, *8*(1), 609. https://doi.org/10.1038/s41467-017-00731-0.

Acemoglu, D., & Jackson, M. O. (2015). History, expectations, and leadership in the evolution of social norms. *Review of Economic Studies*, *82*(2), 423–456.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715–753. https://doi.org/10.1162/003355300554881.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, *128*(2), 469–530.

Anderson, S. P., Goeree, J. K., & Holt, C. A. (2001). Minimum-effort coordination games: Stochastic potential and logit equilibrium. *Games and Economic Behavior*, *34*(2), 177–199. https://doi.org/10.1006/game.2000.0800.

Andreoni, J., Nikiforakis, N., & Siegenthaler, S. (2017). *Social change and the conformity trap*. Working paper.

Arrow, K. (1970). Political and economic evaluation of social effects and externalities. In *The analysis of public output*, NBER Chapters, National Bureau of Economic Research, Inc (pp. 1–30).

Asch, S. E. (1952). Effects of group pressure on the modification and distortion of judgements. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (2nd ed., pp. 2–11). New York: NY Holt.

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

Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. New York: Oxford University Press.

Bond, R. (2005). Group size and conformity. *Group Processes and Intergroup Relations*, *8*(4), 331–354.

Brandts, J., & Cooper, D. J. (2006). A change would do you good... an experimental study on how to overcome coordination failure in organizations. *American Economic Review*, *96*(3), 669–693. https://doi.org/10.1257/aer.96.3.669.

Brandts, J., & Cooper, D. J. (2007). It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association*, *5*(6), 1223–1268. https://doi.org/10.1162/JEEA.2007.5.6.1223.

Brock, W. A., & Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies*, *68*(2), 235–260.

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2018). *Misperceived social norms: Female labor force participation in Saudi Arabia*. NBER Working Papers 24736, National Bureau of Economic Research, Inc.

Cachon, G. P., & Camerer, C. F. (1996). Loss-avoidance and forward induction in experimental coordination games. *The Quarterly Journal of Economics*, *111*(1), 165–194.

Cason, T. N., Savikhin, A. C., & Sheremeta, R. M. (2012). Behavioral spillovers in coordination games. *European Economic Review*, *56*(2), 233–245.

Charness, G., Rigotti, L., & Rustichini, A. (2007). Individual behavior and group membership. *American Economic Review*, *97*(4), 1340–1352. https://doi.org/10.1257/aer.97.4.1340.

Chaudhuri, A., Schotter, A., & Sopher, B. (2009). Talking ourselves to efficiency: Coordination in intergenerational minimum effort games with private, almost common and common knowledge of advice. *Economic Journal*, *119*(534), 91–122.

Chen, R., & Chen, Y. (2011). The potential of social identity for equilibrium selection. *The American Economic Review*, *101*(6), 2562–2589.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *The American Economic Review*, *99*(1), 431–457.

Chernoff, R. A., & Davison, G. C. (2005). An evaluation of a brief hiv/aids prevention intervention for college students using normative feedback and goal setting. *AIDS Education and Prevention*, *17*(2), 91–104.

Cheung, Y. W., & Friedman, D. (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior*, *19*(1), 46–76. https://doi.org/10.1006/game.1997.0544.

Choi, S., & Lee, J. (2014). Communication, coordination and networks. *Journal of the European Economic Association*, *12*(1), 223–247. https://doi.org/10.1111/jeea.12058.

Devetag, G., & Ortmann, A. (2007). When and why? A critical survey on coordination failure in the laboratory. *Experimental Economics*, *10*(3), 331–344. https://doi.org/10.1007/s10683-007-9178-9.

Duffy, J., & Lafky, J. (2018). *Living a lie: Theory and evidence on public preference falsification*. Working paper

Eckel, C., Grossman, P. J., & Milano, A. (2007). Is more information always better? An experimental study of charitable giving and Hurricane Katrina. *Southern Economic Journal*, *74*(2), 388–411.

Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives*, *3*(4), 99–117.

Elster, J. (1990). Norms of revenge. *Ethics*, *100*(4), 862–885.

Fernández, R., & Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, *1*(1), 146–177. https://doi.org/10.1257/mac.1.1.146.

Fershtman, C., & Gneezy, U. (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, *116*(1), 351–377. https://doi.org/10.1162/003355301556338.

Gage, A. J. (1998). Sexual activity and contraceptive use: The components of the decisionmaking process. *Studies in Family Planning*, *29*(2), 154–166.

Galor, O., & Weil, D. N. (1996). The gender gap, fertility, and growth. *The American Economic Review*, *86*(3), 374–387.

Gneezy, U., Leonard, K., & List, J. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, *77*(5), 1637–1664.
Goldin, C. (1986). *The female labor force and American Economic Growth, 1890–1980* (pp. 557–604). Chicago, IL: University of Chicago Press.
Greenwood, J., Seshadri, A., & Yorukoglu, M. (2005). Engines of liberation. *Review of Economic Studies*, *72*(1), 109–133.
Grutzpalk, J. (2002). Blood feud and modernity: Max webers and Émile durkheims theories. *Journal of Classical Sociology*, *2*(2), 115–134. https://doi.org/10.1177/1468795X02002002854.
Hargreaves Heap, S., & Zizzo, D. (2009). The value of groups. *American Economic Review*, *99*(1), 295–323.
Harsanyi, J., & Selten, R. (1988). *A general theory of equilibrium selection in games* (1st ed., Vol. 1). Cambridge, MA: The MIT Press.
Hart, P. S., & Nisbet, E. C. (2011). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research*. https://doi.org/10.1177/0093650211416646.
Hechter, M., & Opp, K. (2001). *Social Norms*. New York: Russell Sage Foundation.
İçli, T. G. (1994). Blood feud in turkey: A sociological analysis. *The British Journal of Criminology*, *34*(1), 69–74. https://doi.org/10.1093/oxfordjournals.bjc.a048384.
Inzlicht, M., & Kang, S. K. (2010). Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology*, *99*(3), 467–481.
Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics*, *7*(1), 63–88. https://doi.org/10.1146/annurev-economics-080614-115404.
Katz, D., & Allport, F. H. (1931). *Students' Attitudes*. Burlington, IA: Craftsman Press.
Keser, C., Suleymanova, I., & Wey, C. (2012). Technology adoption in markets with network effects: Theory and experimental evidence. *Information Economics and Policy*, *24*(3), 262–276. https://doi.org/10.1016/j.infoecopol.2012.03.001.
Knez, M., & Camerer, C. (1994). Creating expectational assets in the laboratory: Coordination in weakest-link games. *Strategic Management Journal*, *15*, 101–119.
Kopanyi-Peuker, A., Offerman, T., & Sloof, R. (2015). *Probation or promotion? The fear of exclusion improves team-production*. Discussion Paper, University of Amsterdam.
Lim, M. (2012). Clicks, cabs, and coffee houses: Social media and oppositional movements in Egypt, 2004–2011. *Journal of Communication*, *62*(2), 231–248. https://doi.org/10.1111/j.1460-2466.2012.01628.x.
Lim, W., & Neary, P. R. (2016). An experimental investigation of stochastic adjustment dynamics. *Games and Economic Behavior*, *100*, 208–219. https://doi.org/10.1016/j.geb.2016.09.010.
Loiseau, E., & Nowacka, K. (2015). *Can social media effectively include womens voices in decision-making processes?*. Issues papers, OECD.
Luhan, W. J., Kocher, M. G., & Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics*, *12*(1), 26–41.
Mann, L. (1977). The effect of stimulus queues on queue-joining behavior. *Journal of Personality and Social Psychology*, *35*(6), 437–442.
Mannix, E. A., & Loewenstein, G. F. (1994). The effects of interfirm mobility and individual versus group decision making on managerial time horizons. *Organizational Behavior and Human Decision Processes*, *59*(3), 371–390.
Michaeli, M., & Spiro, D. (2017). From peer pressure to biased norms. *American Economic Journal: Microeconomics*, *9*(1), 152–216. https://doi.org/10.1257/mic.20150151.
Miller, D. T., & McFarland, C. (1987). Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and Social Psychology*, *53*(2), 298–305. https://doi.org/10.1037/0022-3514.53.2.298.
Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5.
Ochs, J. (2008). Coordination problems and communication. In S. N. Durlauf & L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Basingstoke: Palgrave Macmillan.
Offerman, T., Sonnemans, J., & Schram, A. (2001). Expectation formation in step-level public good games. *Economic Inquiry*, *39*(2), 250–269.
Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, *14*(3), 137–158. https://doi.org/10.1257/jep.14.3.137.

Prentice, DA. (2007). Pluralistic ignorance. In *Encyclopedia of social psychology*. NBER Chapters (vol 1, pp. 674–674). SAGE Publications, Inc. https://doi.org/10.4135/9781412956253

Prentice, D. A., & Miller, D. T. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 28, pp. 161–209). Cambridge, MA: Academic Press. https://doi.org/10.1016/S0065-2601(08)60238-5.

Riedl, A., Rohde, I. M. T., & Strobel, M. (2015). Efficient coordination in weakest-link games. *The Review of Economic Studies*, *83*(2), 737–767. https://doi.org/10.1093/restud/rdv040.

Rothenhäusler, D., Schweizer, N., & Szech, N. (2015). *Institutions, shared guilt, and moral transgression*. CESifo Working Paper 5525, Munich

Sandstrom, M., Makover, H., & Bartini, M. (2013). Social context of bullying: Do misperceptions of group norms influence children's responses to witnessed episodes? *Social Influence*, *8*(2–3), 196–215. https://doi.org/10.1080/15534510.2011.651302.

Schroeder, C. M., & Prentice, D. A. (1998). Exposing pluralistic ignorance to reduce alcohol use among college students. *Journal of Applied Social Psychology*, *28*(23), 2150–2180. https://doi.org/10.1111/j.1559-1816.1998.tb01365.x.

Sherif, M. (1936). *The psychology of social norms*. Manhattan, NY: Harper and Brothers.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Boston, MA: Brooks/Cole.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In W. G. Austin & S. Worchel (Eds.), *Psychology of intergroup relations* (pp. 7–24). Wokingham: Nelson-Hall.

Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, *80*(1), 234–248.

Weber, R. A. (2006). Managing growth to achieve efficient coordination in large groups. *American Economic Review*, *96*(1), 114–126.

Wenzel, M. (2005). Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, *26*(6), 862–883.

Wilder, D. A. (1977). Perception of groups, size of opposition, and social influence. *Journal of Experimental Social Psychology*, *13*(3), 253–268. https://doi.org/10.1016/0022-1031(77)90047-6.