Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-023-01745-4 OPEN

Disentangling material, social, and cognitive determinants of human behavior and beliefs

Denis Tverskoi[1,9], Andrea Guido[2,3,9], Giulia Andrighetto[2,4,5], Angel Sánchez[6,7] & Sergey Gavrilets[1,8]✉

In social interactions, human decision-making, attitudes, and beliefs about others coevolve. Their dynamics are affected by cost-benefit considerations, cognitive processes (such as cognitive dissonance, social projecting, and logic constraints), and social influences by peers (via descriptive and injunctive social norms) and by authorities (e.g., educational, cultural, religious, political, administrative, individual or group, real or fictitious). Here we attempt to disentangle some of this complexity by using an integrative mathematical modeling and a 35-day online behavioral experiment. We utilize data from a Common Pool Resources experiment with or without messaging promoting a group-beneficial level of resource extraction. We directly estimate the weights of different factors in decision-making and beliefs dynamics. We show that personal norms and conformity with expected peers' actions have the largest impact on decision-making while material benefits and normative expectations have smaller effects. Individuals behaving prosocially are characterized by higher weights of personal norms while antisocial types are more affected by conformity. Messaging greatly decreases the weight of personal norms while simultaneously increases the weight of conformity. It also markedly influences personal norms and normative expectations. Both cognitive and social factors are important in the dynamics of beliefs. Between-individual variation is present in all measured characteristics and notably impacts observed group behavior. At the same time, gender differences are small. We argue that one can hardly understand social behavior without understanding the dynamics of personal beliefs and beliefs about others and that cognitive, social, and material factors all play important roles in these processes. Our results have implications for understanding and predicting social processes triggered by certain shocks (e.g., social unrest, a pandemic, or a natural disaster) and for designing policy interventions aiming to change behavior (e.g., actions aimed at environment protection or climate change mitigation).

[1] Center for the Dynamics of Social Complexity, National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA. [2] Institute of Cognitive Sciences and Technologies, Italian National Research Council, Rome, Italy. [3] CEREN EA 7477, Burgundy School of Business, Université Bourgogne Franche-Comté, Dijon, France. [4] Institute for Futures Studies, Stockholm, Sweden. [5] Malardalens University, Vasteras, Sweden. [6] Grupo Interdisciplinar de Sistemas Complejos, Departamento de Matemáticas Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain. [7] Instituto de Biocomputación y Física de Sistemas Complejos, Universidad de Zaragoza, 50018 Zaragoza, Spain. [8] Department of Ecology and Evolutionary Biology, Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA. [9] These authors contributed equally: Denis Tverskoi, Andrea Guido. ✉email: gavrila@utk.edu

Introduction

A complex web of material, social, and cognitive factors and forces influence the behavior and beliefs of individuals and groups engaged in social interactions. The question of their relative importance has been controversial for centuries. Arguments emphasizing material cost-benefit considerations have been at the center of many philosophical theories aiming to explain human nature (Marx 1959). These arguments are formalized in various flavors of game theory—classical (Fudenberg and Tirole 1992), evolutionary (Sandholm, 2010), mean-field (Tembine 2017), and quantum (Piotrowski and Sladkowski 2003). That humans are deeply social was already clear to Aristotle and many influential philosophers after him. Various mathematical models of social learning, imitation, and opinion spread capture this human feature (DeGroot 1974, Friedkin et al. 2016, Galesic and Stein 2019, Rashevsky 1949, Watts 2002) We also care about how our actions are perceived by peers (Bicchieri 2006, Cialdini et al. 1990), and there are mathematical approaches describing the effects of social norms, namely the unwritten social rules, on behavior (Gavrilets 2020, Young 2015). Moreover, humans are greatly influenced by authorities (e.g., educational, cultural, religious, political, administrative, individual or group, real or fictitious (Bernays 1928)). Furthermore, our actions and beliefs are affected by psychological processes. Cognitive dissonance (i.e., a feeling of mental discomfort experienced when the person's attitudes, beliefs, or behaviors conflict) can cause changes in behaviors but also in attitudes or beliefs (Festinger 1957). To predict the intentions and beliefs of others, we may use the "theory of mind" (Premack and Woodruff 1979) and social projection, which is a tendency to assume that others are similar to oneself (Krueger 2007). However, across societies, individuals widely misperceive what others think and what others do (Bursztyn and Yang 2021). Logic constraints (Friedkin et al. 2016, Rawlings 2020) can put certain bounds on what we can believe simultaneously. We also make errors.

Here we attempt to disentangle some of this complexity by using a novel integrative mathematical model (Gavrilets 2021) and a behavioral experiment designed to validate and parameterize the model. Whereas earlier work has mostly focused on one or two of the variety of factors discussed above at a time, we bring together a large number of those and contrast their effects on human actions, attitudes and beliefs about others. When modeling individual actions, we explicitly account for the effects of expected material payoffs, cognitive dissonance, and three different types of social influence (expected disapproval by peers and authority and conformity with peers' actions) in the utility function. Because actions of individuals depend on their changing beliefs, we also explicitly model belief dynamics as driven by certain cognitive processes and learning from observed actions of others, closing the feedback loop.

Earlier experimental work on conditional cooperation has demonstrated that individual behavior is influenced by beliefs about the behavior of others (Andreozzi et al. 2020, Fischbacher and Gächter 2010), that there are multiple channels through which personal beliefs and social influence can affect economic behavior (Basić and Verrina 2020, Kölle and Quercia 2021), and that there is substantial variation in the effectiveness of different behavioral interventions (Mertens et al. 2022, Sunstein 2021, Thaler and Sunstein 2021). Here we advance this line of research by using our integrative model to study the co-evolution of decision-making with attitudes and beliefs in situations in which individuals receive or not receive a message from an authority informing them how to act to maximize the total group benefit. To this aim, we conduct a long-term (35 days) behavioral experiment where we measured not only the actions of participants, but we also elicited their personal norms and beliefs about others, using an incentivized procedure, and tracked their evolution over time. The model validation involved its comparison with a variety of other models in the ability to describe the social dynamics observed in our experiments. Overall our modeling and experiment provide a very coherent picture of the way decisions are made and personal attitudes and beliefs about others change. Our approach allows us not only to evaluate the statistical significance of different forces driving these changes but also to compare directly their effects in decision-making and belief dynamics.

Model

Consider individuals interacting in groups. Let a continuous variable x specify an action chosen by a focal individual. Each individual is characterized by an attitude y which specifies the most appropriate action in a given social situation as perceived by the individual. Each individual also has a belief (prediction) $\tilde{x}$ about the average action of peers as well as a second-order belief $\tilde{y}$ about the average attitude of their peers. In the social psychology terminology which we will use, variables $y, \tilde{x}$, and $\tilde{y}$ are called a personal norm, an empirical expectation, and a normative expectation, respectively (Bicchieri 2006, Cialdini et al. 1990, Schwartz 1977, Szekely et al. 2021). The empirical expectation $\tilde{x}$ can also be viewed as a descriptive norm (i.e., the most common behavior) while the normative expectation $\tilde{y}$ as an injunctive norm (i.e., the socially appropriate behavior) (Bicchieri 2006, Cialdini et al. 1990, Gavrilets 2020), both as perceived by the individual. In our experiment to be described below we measured all these variables directly. Individuals are also subject to influence by an external authority promoting a particular action G. We assume $x, y, \tilde{x}, \tilde{y}, G$ are non-negative.

Following ref. Gavrilets (2021), we postulate that each individual chooses an action x in an attempt to maximize the subjective utility function

$$u = \underbrace{A_0\pi(x,\tilde{x})}_{\text{material payoff}} - \underbrace{\frac{1}{2}A_1(x-y)^2}_{\text{cognitive dissonance}} - \underbrace{\frac{1}{2}A_2(x-\tilde{y})^2}_{\text{disapproval by peers}} - \underbrace{\frac{1}{2}A_3(x-\tilde{x})^2}_{\text{conformity w/ peers}} - \underbrace{\frac{1}{2}A_4(x-G)^2}_{\text{compliance w/ authority}}. \quad (1)$$

That is, individuals expect to get a material payoff $\pi(x, \tilde{x})$ which depends on the expected action $\tilde{x}$ of their groupmates. They also pay psychological costs if their action x deviates from what they believe is the right action (y) due to cognitive dissonance (Rabin 1994), from what they think their peers and the authority expect from them ($\tilde{y}$ and G, respectively), and also by not conforming with the expected average behavior of their group ($\tilde{x}$). Non-negative constant parameters $A_0,\ldots, A_4$ measure the weights of the corresponding terms in the utility function. The utility function (1) was introduced as a generalization of utility functions in earlier work which included the terms accounting for material payoffs, cognitive dissonance, and conformity (Akerlof and Dickens 1982, Calabuig et al. 2018, Kuran and Sandholm 2008, Rabin 1994).

Assume that the partial derivative $\frac{\partial \pi(x,\tilde{x})}{\partial x}$ is a linear function of its arguments. Let θ be the action maximizing the expected material payoff $\pi(x, \tilde{x})$; θ can be found in a straightforward way (see Supplementary Materials, SM). Given the utility function (1), the best response action can be written as a weighted sum of the values maximizing the corresponding components in the utility function

$$x = \max(0, B_0\theta + B_1 y + B_2\tilde{y} + B_3\tilde{x} + B_4 G). \quad (2)$$

Coefficients B_i are the relative weights of material factors, personal norms, normative expectations, empirical expectations, and the authority's messaging in the decision made, respectively ($\sum B_i = 1$). They are directly related to coefficients A_i of the utility function (SM).

After taking actions and observing behavior of groupmates, the attitude and beliefs of a focal individual change. We describe these changes using linear recurrence equations:

$$y' = y + \underbrace{\alpha_1(x-y)}_{\text{cognitive dissonance}} + \underbrace{\beta_1(X-y)}_{\text{conformity w/ peers}} + \underbrace{\gamma_1(G-y),}_{\text{compliance w/ authority}} \quad (3a)$$

$$\tilde{y}' = \tilde{y} + \underbrace{\alpha_2(y-\tilde{y})}_{\text{social projection}} + \underbrace{\beta_2(X-\tilde{y}),}_{\text{learning about others}} + \underbrace{\gamma_2(G-\tilde{y}),}_{\text{compliance w/ authority}} \quad (3b)$$

$$\tilde{x}' = \tilde{x} + \underbrace{\alpha_3(\tilde{y}-\tilde{x})}_{\text{logic constraints}} + \underbrace{\beta_3(X-\tilde{x}),}_{\text{learning about others}} + \underbrace{\gamma_3(G-\tilde{x}),}_{\text{compliance w/ authority}} \quad (3c)$$

where the prime means the next time step, X is the average action of groupmates as observed by the focal individual (so that different individuals can have different X), and $\alpha_i, \beta_i, \gamma_i$ are non-negative constant coefficients measuring the strength of the corresponding forces. Here the "cognitive dissonance" term in equation (3a) acts to change individual attitude y to justify the action x previously chosen. The "social projection" term in equation (3b) captures the ego's belief that others are probably similar to themselves (Krueger 2007, Premack and Woodruff 1979). The "logic constraints" term in equation (3c) reduces the mismatch between the ego's beliefs about actions and beliefs of others (cf., ref. (Friedkin et al. 2016)). The "conformity w/ peers" and the two "learning about others" terms move the corresponding beliefs closer to the observed average behavior X of peers (Kashima et al. 2015). The "compliance w/ authority" terms move the corresponding beliefs closer to the promoted "standard" G. The authority's messaging effectively changes the utility function (1) and simultaneously affects beliefs (equations 3) which then feed back into the utility function and behavior. All parameters defined above are individual-specific.

The beliefs dynamics equations 3 were introduced in ref. (Gavrilets 2021) as a simple generalization of earlier models focusing on the changes in personal attitudes (or opinions) y as a result of the exchange of opinions between individuals (Centola et al. 2005, DeGroot 1974, Friedkin et al. 2016, Galesic and Stein 2019, Gavrilets et al. 2016, Gavrilets 2003, Kashima et al. 2021, Redner 2019, Watts 2002). Such changes are usually described by equations analogous to the second terms in equation (3a). Similar linear equations are also used in cognitive neuroscience (Olsson et al. 2020). ref. Gavrilets (2021) used the same logic to also describe the changes in normative $\tilde{y}$ and empirical $\tilde{x}$ expectations (equations (3b) and (3c)) as well as to capture the influence of authority's messaging (Gavrilets 2003). The latter effect is described by the last term in each of the equations 3.

We stress that all socio-psychological factors included in our model have been repeatedly shown to be important in decision-making and belief dynamics (see the references above). The specific functions describing them were chosen because of their intuitive nature, mathematical simplicity, existing tradition, generalizability, and the easiness of statistical evaluation (see below). We also note that our modeling answers recent calls to include individual beliefs in utility functions used in economics (Loewenstein and Molnar 2018, Molnar and Loewenstein 2022) and to integrate social and cognitive aspects of belief dynamics (Galesic et al. 2020).

A Common Pool Resources experiment

Gavrilets (2021) focused on the theoretical properties of equilibria in this model. Here we aim to validate and parameterize it. Specifically, we conducted a long-term, online experiment implementing a non-linear Common Pool Resources (CPR) game which is often viewed as "a far more realistic environment ... than many of the [other] dilemma games" (Ostrom et al. 1992). Examples include fish in bodies of water, pastures, and water used for irrigation. These games are commonly used in experimental studies of social dilemmas (Apesteguia 2006, Apesteguia and Maier-Rigaud 2006, Ostrom et al. 1992, Walker et al. 1990) because of their realism.

More formally, we assume each individual in a group of size n has an endowment of resources π_0 which can be invested to extract resources from the CPR or in a safe, outside activity. Individuals independently decide on their own level of investment x_i in the extraction of the CPR. The group investment, i.e., the sum of individual investments, $Z = \sum x_i$ defines the collective return P. It is usually assumed that P is a quadratic function $P = bZ - 0.5dZ^2$, where b and d are constant positive parameters. The share of the produced resource P that the individual receives is equal to their share of the group investment: $v_i = x_i/Z$. We write the individual payoff as $\pi_i = \pi_0 + v_i P - cx_i$, where c is the cost coefficient associated with extraction effort. Due to the quadratic form of the group return function P, it is more beneficial to invest in the extraction of the CPR relative to the safe investment if the group's total investment Z is small. However, as Z increases, investing in the CPR extraction becomes counterproductive. In this model, the self-interested investment (Nash equilibrium) is $x_{\mathrm{NE}} = \frac{2(b-c)}{d(n+1)}$, the investment maximizing group benefit (Pareto efficient equilibrium) is $x_{\mathrm{opt}} = \frac{b-c}{nd}$, and the value maximizing payoff given an empirical expectation $\tilde{x}$ is $\theta = \max\left(0, \frac{b-c}{d} - \frac{n-1}{2}\tilde{x}\right)$ (see SM). Note that x_{NE} is larger than x_{opt} so self-interested individuals are predicted to over-exploit the resource suffering a payoff loss. Gavrilets (2021) showed that in this game adding an external authority promoting a socially optimal individual effort can backfire, that is, individuals can respond by increasing rather than decreasing their efforts.

The experiments took place in the spring of 2021 and lasted 35 rounds (1 round per day). We followed the general experimental protocol of (Szekely et al. 2021) (see SM). For each round, the subjects were randomly reshuffled into groups of size $n = 6$ and provided with an endowment of $\pi_0 = 30$ points each. The investment x devoted to the CPR extraction was described to the subjects as a contribution to the "Common Account" while the points not invested in resource extraction as a contribution to a "Personal Account" (see SM).

Each round, we first elicited their personal norms y, normative expectations $\tilde{y}$, and empirical expectations $\tilde{x}$ regarding individual investments in the CPR extraction of other individuals. Specifically, subjects were asked to answer three questions: 1). "In your opinion, how many points should a participant from your group, including yourself, put into the Common Account in this round?" This was our variable y. 2). "How many points will the other five participants in your group put into the Common Account in this round?" The mean of these numbers was our variable $\tilde{x}$. 3). "How many points do the other five participants in your group think that you should put into the Common Account in this round?" The mean of these numbers was our variable $\tilde{y}$. After answering these questions the subjects made their decision on their own investment x. The questions about normative and empirical

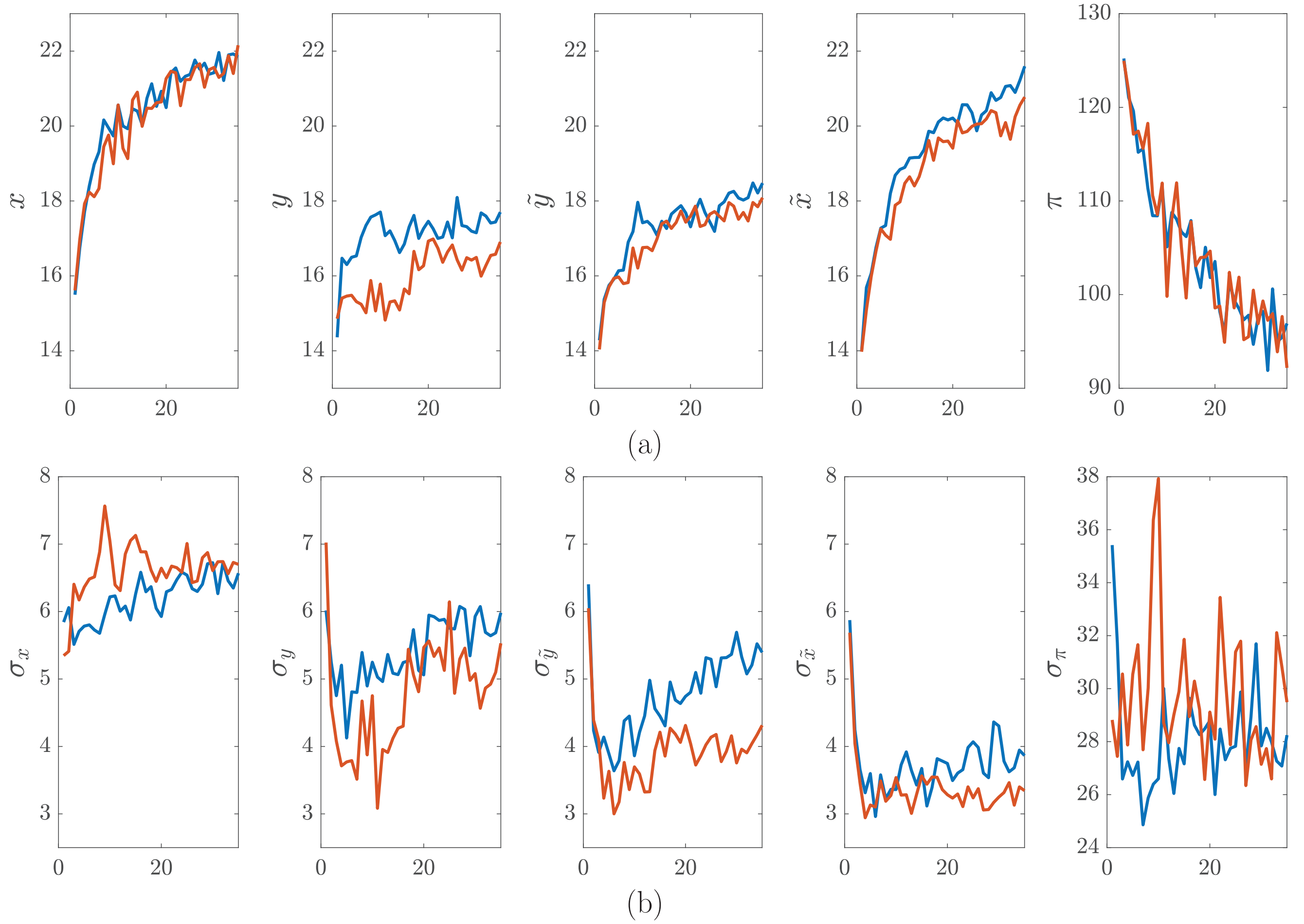

Fig. 1 The dynamics of means and standard deviations of the main variables in the experiment with (red) and without (blue) messaging. a Dynamics of mean values. **b** Dynamics of standard deviations.

expectations were incentivized (Bicchieri 2006, Szekely et al. 2021) by paying an amount of money dependent on the accuracy of their predictions.

After each round, the subjects were informed about their own payoffs and the actions taken by their groupmates. Subjects were provided with a tool to evaluate their payoffs given their and others' actions. In our experiment, we set $b = 15, d = 1/6, c = 1$, so that the symmetric Nash equilibrium was $x_{\mathrm{NE}} = 24$ points while the investment maximizing the group's benefit was $x_{\mathrm{opt}} = 14$ points. We recruited 300 subjects who were equally divided between two treatments: without and with messaging. In the experiment with messaging, at each round subjects saw a message "Please note that the total group profit is maximized if each player contributes 14 points to the Common Account." This treatment is aimed to test the effects of authority messaging on actions and beliefs dynamics (Croson and Marks 2001, Dal Bó and Dal Bó 2014).

Besides participating in the CPR experiments, each subject responded to the Big Five questionnaire (Benet-Martínez and John 1998, John et al. 1991), Risk Preferences task (Dave et al. 2010), Rule Compliance task (Kimbrough and Vostroknutov 2018), a Demographic questionnaire and Social Value Orientation task (Murphy et al. 2011).

Results

Figure 1 illustrates the observed dynamics of our focal variables. There are several interesting observations. The average individual extraction effort x appears to evolve to a value around 22 which is below the Nash equilibrium, $x_{\mathrm{NE}} = 24$ (see Figs. S2, S3 in SM). Subjects extract more from the CPR than what they believe is right (i.e., $x > y$), and they also expect others to do the same ($\tilde{x} > y$). They extract more than what they think others will (i.e., $x > \tilde{x}$). Personal norms y equilibrate faster than other variables (see SM). As predicted in Gavrilets (2021), variation between individuals is the highest in individual actions, followed by that in personal norms, normative expectations, and empirical expectations, respectively. Interestingly, messaging does not affect the average extraction efforts and payoff, but it increases their standard deviations. (This observation is discussed below.) It also decreases $y, \tilde{y}, \tilde{x}$ (i.e., moves them closer to the value promoted by messaging) and their variation. The average personal norm evolves to a value above the socially optimal extraction level of 14 promoted by messaging (see SM). The average payoff continuously declines over time.

To conduct a more in-depth analysis of our data, we needed a statistical method capable of simultaneously addressing significant between-individual differences (necessitating the use of models with heterogeneous slopes), parameter constraints, and multicollinearity. We found that panel data modeling, mixed model approaches, and Markov Chain Monte Carlo (MCMC) techniques were less efficient in this context (as detailed in section 4.1 of SM) compared to the mean group estimation method (Pesaran and Smith 1995). This latter approach involves estimating the parameters of regressions ((2), 3) individually for each subject before averaging them across the entire group. Not only is this method commonly employed for data collected over time and across heterogeneous units (Adeneye et al. 2021, Dong et al. 2017, Lee and Sul 2022, Marlowe 2004, Ndambendia and Njoupouognigni 2010, Paramati and Roca 2019, Sadorsky 2013, Teal and Eberhardt 2010), but it is also computationally simple and offers a direct means of addressing constraints and multicollinearity. Our simulations (discussed below) demonstrate the method's ability to recover known parameter values from simulated data and accurately describe observed mean trajectories. Further discussion and justification of our chosen methodology can be found in SM.

For each individual, we estimated six parameters for actions (B_0, B_1, B_2, B_3, intercept C and error σ), and four (without messaging) or five (with messaging) parameters for each of the beliefs ($\alpha_i, \beta_i, \gamma_i$, intercept C, error σ). The intercepts were included to account for the effects of factors not captured by our model. Using our data we estimated 13 out of 14 parameters of our model. We could not estimate parameter B_4. In the experiments, we used a single value of $G = 14$ implying that the term B_4G in the best response equation (2) is a constant which cannot be differentiated from the effects of other forces captured by the intercept C. For each individual, we compared 32 different models for actions and 8 (without messaging) or 16 (with messaging) models for each of the beliefs. For example, in the case of actions x, there are four independent variables potentially affecting behavior: $\theta, y, \tilde{y}, \tilde{x}$ plus an intercept. This leads to $2^5 = 32$ different combinations of factors included in each model. For each model, we first tested for multicollinearity using condition numbers and variance decompositions (Belsley 1991, Belsley et al. 2005). If multicollinearity was detected, we applied the ridge regression technique (Hoerl and Kennard 1970). We compared candidate models using AICc which is a modification of Akaike Information Criterion (AIC) for small sample size (Burnham and Anderson 2001). The differences in the AICc values between some models were often small. Therefore, rather than choosing parameters of the single best model on the basis of AICs, we used the standard method of model averaging, weighting model estimates with AIC weights (Burnham and Anderson 2001). To find confidence intervals of our mean group estimates we used nonparametric bootstrap. We tested our approach using artificial data generated by our dynamic model specified by equations ((2), 3). (see the SM).

Parameter estimates. Figure 2a illustrates the average estimates of parameters of the best response function. Without messaging, cognitive dissonance, which forces individuals to align their actions with personal norms, has the largest weight ($B_1 = 0.32$) in decision-making. In our framework, a personal norm is a preferred action in the absence of any other material or non-material influences; it could be based on internalized values (Schwartz 1977) or some simple heuristic rules (Capraro et al. 2014). Conformity with peers' actions has the second largest weight $B_3 = 0.16$. When messaging is introduced, it significantly reduces the weight of personal norms to $B_1 = 0.19$ simultaneously increasing the weight of conformity to $B_3 = 0.21$. It thus appears that messaging removes to a certain extent the need to perform complex cognitive evaluations required for deciding what is the right action to take while at the same time causes individuals to pay closer attention to whether others comply with the authority messaging. The weights of material factors and normative expectations are the smallest and are not affected much by messaging ($B_0 = 0.11, B_2 = 0.09$ without messaging and $B_0 = 0.12, B_2 = 0.10$ with messaging). The low weight of material payoffs is surprising. The small effect of normative expectations (B_2) is likely because individuals were randomly reshuffled every day between groups and the need for cooperation (i.e., limited extraction of the CPR) was not made salient to the subjects.

Figure 2b illustrates the average estimates of parameters controlling belief dynamics. It shows that the effects of cognitive factors (cognitive dissonance α_1, social projection α_2, and logic constraint α_3) and those of learning about peers ($\beta_1, \beta_2, \beta_3$) are comparable in magnitude except for the case of empirical expectations where learning about peers is much more important. The last observation is intuitive. The effect of authority's messaging (measured by parameters γ_i) dominates other factors for personal norms (when its weight is more than three times that of the joint effect of conformity and cognitive forces) and for normative expectations (when its weight is about 1.5 times that of the joint effect of conformity and cognitive forces). The effect of messaging is weak for empirical expectations as the subjects can observe their peers' behavior directly. Overall, messaging decreases the effects of cognitive forces (measured by α_i) and the effects of observations on personal norms (measured by β_1) and on normative expectations (measured by β_2).

Heterogeneity. Subjects exhibit high variation in all coefficients we have estimated (see Figs. S9 and S10 in SM). The observed distributions are highly asymmetric. To get better ideas about the effects of between-individual variation, we performed cluster analysis based on estimated coefficients (using k-means and Gaussian mixture models) and on time-series of individual actions x (using the Dynamic Time Warping distance).

The results of k-means clustering are more transparent than those based on Gaussian mixture models so we present only them. Specifically, in terms of parameters of utility function, there are three clusters which are similar between the two experiments (see Fig. 3). Cluster 1 (28% and 16% of subjects in the experiments without and with authority's messaging, respectively) represents individuals who make their decisions mostly on the basis of their personal norms; for them B_1 is the largest parameter. Individuals in this cluster make the smallest investment in the extraction of resources from the CPR. Cluster 2 (15% and 32% of subjects in experiments without and with messaging, respectively) represents individuals who make their decisions mostly on the basis of their empirical expectations/conformity; for them B_3 is the largest parameter. Individuals in this cluster make the largest extraction effort. In cluster 3, which is the largest (57% and 52% of subjects in experiments without and with messaging, respectively), individuals are similarly affected by material, cognitive, and social factors (their coefficients B_i are comparable in magnitude). Figure S12 in the SM illustrates the differences between these clusters in parameters controlling the belief dynamics.

Clustering of parameters of belief dynamics identifies 3 clusters with no messaging and 2 clusters with messaging. With no messaging (see Fig. S15a, b), cluster 1 (23% of subjects) represents individuals whose personal norms do not change much over time (i.e., their α_1, β_1 are very close to zero) and whose normative expectations are mostly defined by personal norms (i.e., α_2 is large). All beliefs of individuals in cluster 2 (30% of subjects) are mostly affected by the actions of others (i.e., they have relatively large coefficients β_i). In cluster 3 (47% of subjects), cognitive and social forces have comparable effects on all beliefs (i.e., α_i and β_i are comparable for all i). The investment in the CPR extraction of individuals in cluster 1 are the smallest while those in cluster 2 are the largest. With messaging, in cluster 1 (57% of subjects) personal norms and normative expectations are determined mostly by messaging (i.e., their γ_1 and γ_2 are large compared to other parameters). In cluster 2 (43% of subjects), all the coefficients α_k, β_k, and γ_k $(k = 1, 2, 3)$ are comparable in

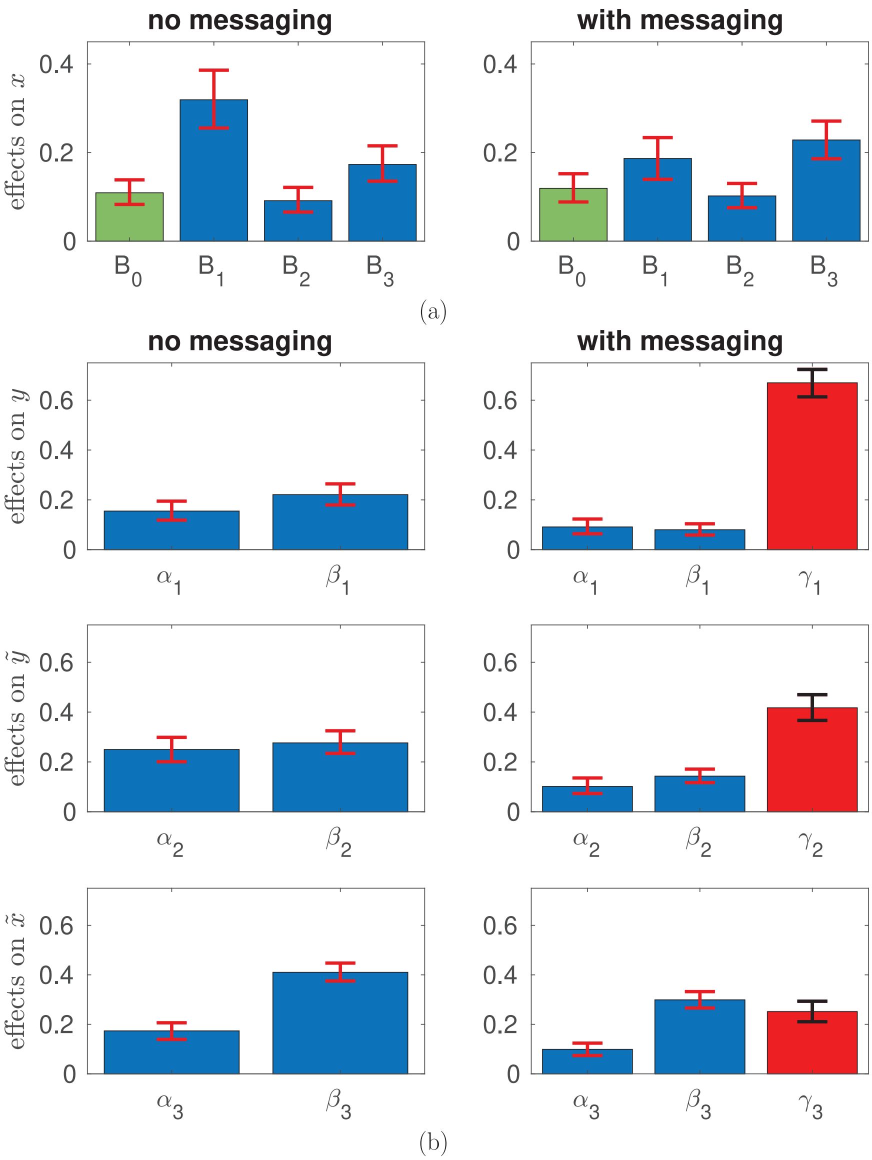

Fig. 2 Average parameter estimates with 95% bootstrap confidence intervals. a Parameters of the best response function (2). **b** Parameters of beliefs dynamics equations 3. Effects of material factors are shown in green; effects of messaging are shown in red.

magnitude. Figure S15 illustrates the differences between these clusters in parameters controlling decision-making.

Clustering based on time-series of actions x identifies two clusters in both cases, with and without messaging (Fig. 4). In cluster 1 (43% of subjects in both experiments), individual variables $x, y, \tilde{y}, \tilde{x}$ and payoffs π are smaller than in cluster 2 (57% of subjects in both experiments). Individuals in cluster 1 have higher weights B_1 of personal norms in their decision-making (see Fig. 4c, d). In the case with messaging, they also have lower weights B_0 of material factors. Without messaging, cluster 1 individuals have lower weight B_3 of empirical expectations/conformity. Figures S17 in SM illustrates the differences between the clusters in parameters controlling beliefs dynamics.

In spite of significant between-subject differences described above, gender differences are small (for more details, see Fig. S18 in SM). Without messaging, males appear to be motivated more by material payoffs while females are more sensitive to personal norms and normative expectations. Males are more affected by authority's messaging.

Prosociality. We used the standard Social Value Orientation task (Murphy and Ackermann 2014, Murphy et al. 2011) to measure the degree of prosociality of our subjects (see SM). Out of 4 possible types (competitive, individualistic, prosocial and altruistic), only a couple were competitive or altruistic types, so we ignored them focusing on individualistic and prosocial subjects. Figures 5 and S19 in the SM show the corresponding average trajectories as well as the estimates of parameters in these two groups. Without messaging, the levels of investment in the CPR extraction x are similar between the two groups. With messaging, prosocial individuals decrease their efforts while individualistic types increase it, presumably to take advantage of the situation. Prosocial individuals have higher weight B_1 of personal norms (significantly higher in the presence of massaging) and weaker reaction to normative and empirical expectations (measured by B_2 and B_3). Prosocial individuals also have higher responsiveness to messaging (γ_i). Messaging decreases the weight B_1 of personal norms and increases the weight B_3 of conformity with peers in both types.

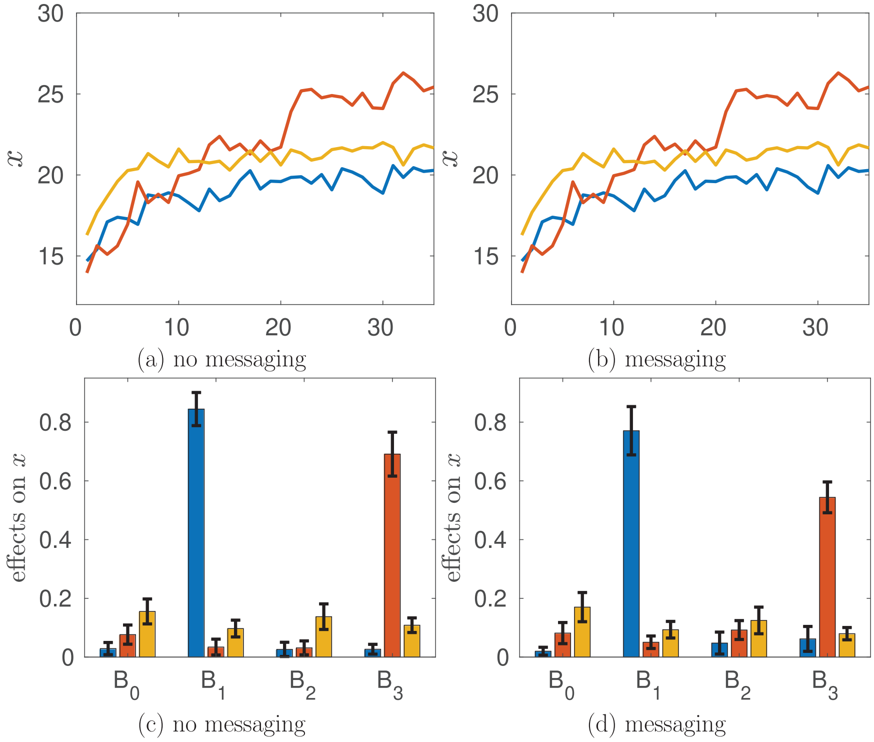

Fig. 3 Differences in the dynamics of actions and parameters of the utility function for individuals from 3 clusters defined on the basis of parameters *Bi* of the utility function. a Mean values of *x* with no messaging. **b** Mean values of *x* with messaging. **c** Mean parameter values with no messaging. **d** Mean parameter values with messaging. Cluster 1: in blue, cluster 2: red, cluster 3: yellow.

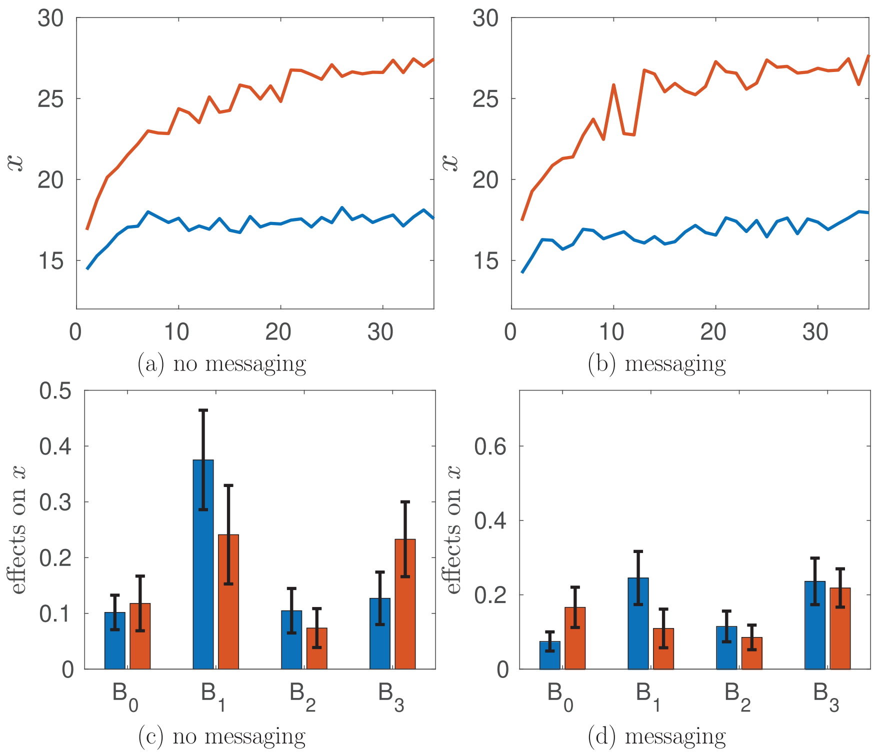

Fig. 4 Differences in the dynamics of actions and parameters of the utility function for individuals from clusters defined on the basis of time series of *x*. a Mean values of *x* with no messaging. **b** Mean values of *x* with messaging. **c** Mean parameter values with no messaging. **d** Mean parameter values with messaging. Characteristics of clusters 1 and 2 are colored in blue and red, respectively.

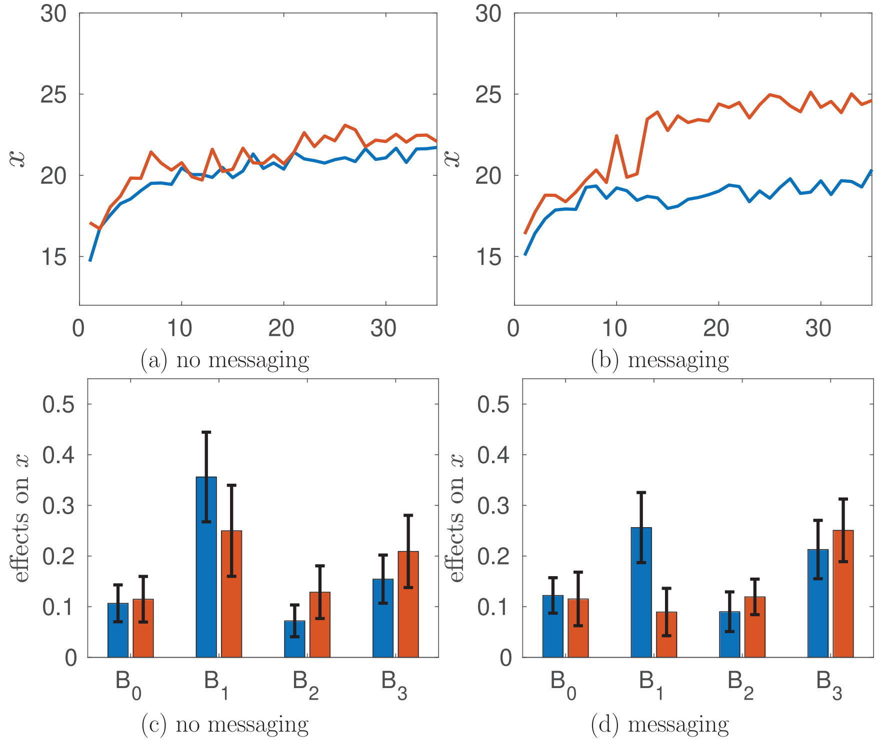

Fig. 5 Differences in the dynamics of actions and parameters of the utility function for individualistic (red) and prosocial (blue) subjects. a Mean values of *x* with no messaging. **b** Mean values of *x* with messaging. **c** Mean parameter values with no messaging. **d** Mean parameter values with messaging.

Rule-following. We also measured the rule-following tendency of our subjects using the Ball task, a method proposed and validated by Kimbrough and Vostroknutov (2016, 2018). Using this method we identified two groups of subjects: with low (≤25%) and high (≥75%) rates of compliance with rules. Figure 6 and S20 in the SM show the corresponding average trajectories as well as the estimates of parameters in these two groups. There is not much difference between the rule-followers and rule-breakers in the absence of messaging. However, with messaging rule-followers decrease their efforts while rule-breakers increase them. This difference also affects their beliefs (see the SM). With messaging, rule-followers have higher weight of personal norms B_1 and responsiveness to messaging γ_i. Messaging decreases the weight of personal norms B_1 and increases the weight of conformity B_3 in both types with rule-breakers showing the largest effect.

Conclusions on heterogeneity. Individuals making larger investments in the CPR extraction have higher payoffs; they are analogous to defectors types in standard social dilemmas. Individuals making smaller extraction efforts can be viewed as cooperators. Our results on clustering of the parameters B_i of the best response function and of time-series of extraction efforts x show that cooperators are mostly affected by their personal norms (their B_1 is the highest) while defectors are more affected by conformity (they have larger values of B_3 except for estimates based on time-series in the case with messaging). Cooperators also have smaller values of B_0 and B_2 (Fig. 3). Our analysis of prosociality and rule-following (illustrated in Figs. 5 and 6) provides further (and independent) evidence that our estimates of parameters capture intended effects. Indeed, prosocial individuals have larger values of B_1 than individualist types while the latter have higher values of B_3. Prosocial individuals also have smaller values of the weight of normative expectations B_2 implying that their reduced efforts are likely motivated not by expected disapproval of others but by internalized values. Messaging has stronger effects on normative and empirical expectations in prosocial types than in individualist types (see Figure S19 in the SM). In the presence of messaging, rule-followers have higher weight of personal norms B_1 but smaller weights of normative B_2 and empirical B_3 expectations which is consistent with behavior of prosocial individuals. Messaging has stronger effects on rule-followers. (For an additional analysis of the differences between pro-social/individualistic and rule-followers/rule-breakers in our experiment see Guido et al. (2023)).

The variation between subjects helps to understand the apparent weakness of the direct effect of messaging on behavior mentioned above (see Fig. 1, top left). What happens is that while prosocial individuals and rule-followers reduce their investment in the CPR extraction in the presence of messaging, individualistic types, and rule-breakers opportunistically increase their investment (see Figs. 5 and 6). These two effects cancel each other with respect to the average effort x but are manifested in the increased variation in x (see Fig. 1).

Predicting trajectories. Figure 7 and S6 in the SM illustrates the predictive ability of our model. In these figures, to obtain "predicted" trajectories we used the obtained parameter estimates and the actual individual data in each round $(\theta, y, \tilde{y}, \tilde{x}, X)$ to predict their values in the next round. The "simulated trajectories" were obtained by repeatedly iterating the dynamic equations ((2), 3) using the obtained estimates of parameters and the actual individual data in the first round. In simulations, we reshuffled individuals between the groups randomly without attempting to recreate the exact history of individual movements between different groups; the results shown are averages over 500 runs. Both for predicted trajectories and for simulated trajectories we used the initial conditions observed in the corresponding experiment. Overall, given all the stochasticity and estimation errors involved, the match between the observed, predicted, and simulated

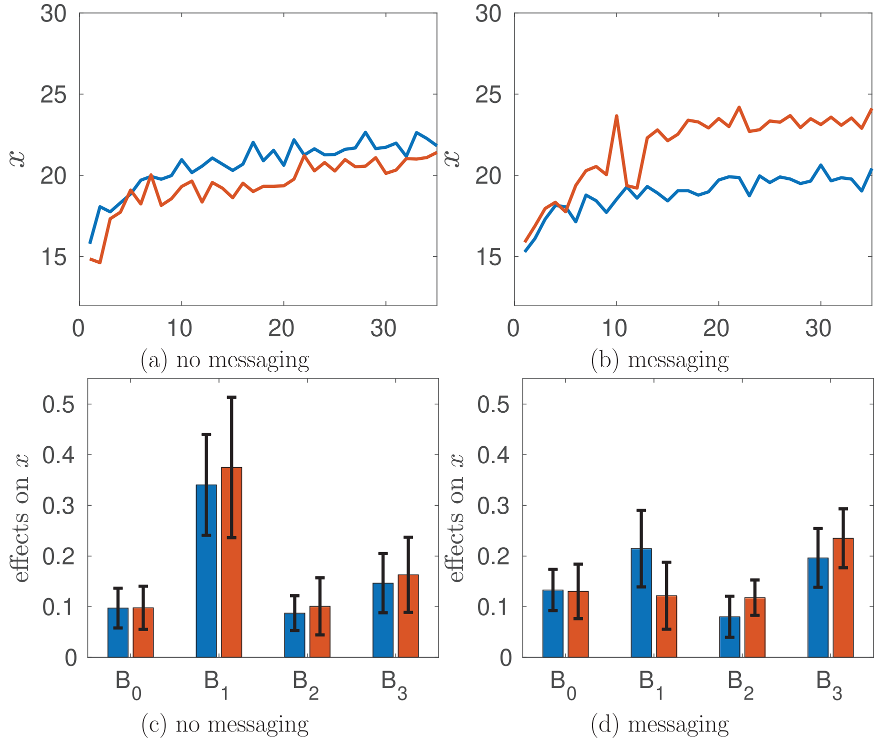

Fig. 6 Differences in the dynamics of actions and parameters of the utility function for rule-breakers (red) and rule-followers (blue). a Mean values of x with no messaging. **b** Mean values of x with messaging. **c** Mean parameter values with no messaging. **d** Mean parameter values with messaging.

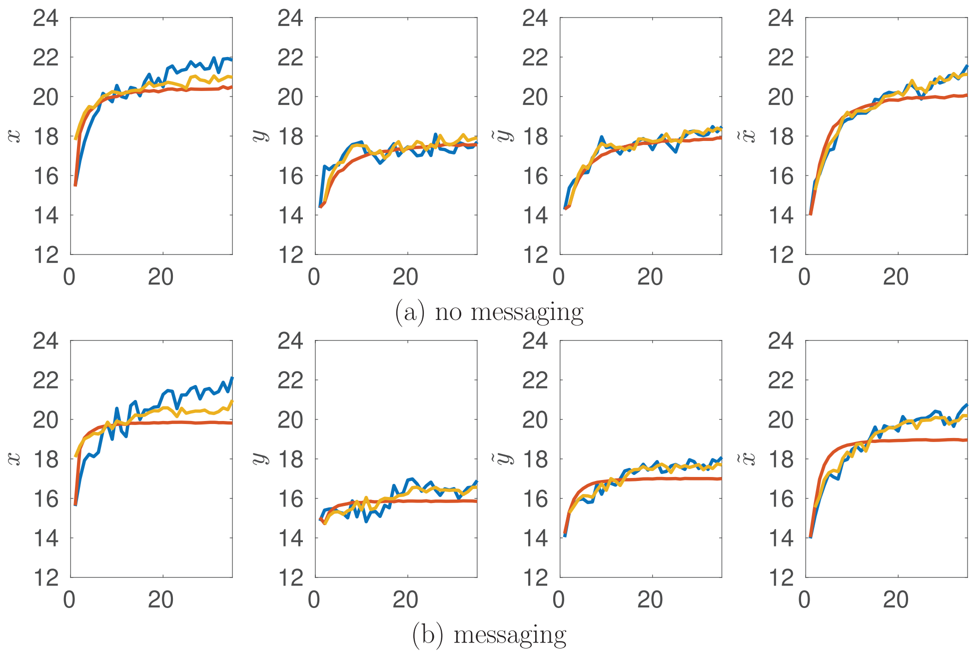

Fig. 7 A comparison between observed (blue), predicted (yellow), and simulated (red) mean trajectories. a No messaging. **b** With messaging.

trajectories is rather good. There is some mismatch for x and $\tilde{x}$, especially in the case of messaging. We explore the causes for this mismatch further in the SM.

Discussion

From our everyday experience we all know that material, social and cognitive factors are important for our actions and beliefs. Research shows that these factors have statistically significant effects on individual behavior and beliefs (Ajzen 1991, Andreozzi et al. 2020, d'Adda et al. 2020, Fishbein and Ajzen 1975, Górges and Nosenzo 2020, Szekely et al. 2021). Here we have developed and empirically tested an integrative approach that allowed us to go beyond identifying statistically significant effects. Specifically, we have directly measured and compared the weights of three different types of factors affecting human behavior and beliefs including 1) expected material payoffs, 2) social influences (by peers and authority), and 3) cognitive forces (such as cognitive dissonance, social projection, logic constants) on individual behavioral actions, personal norms, and normative and empirical expectations. We have done so by using a long-term experiment based on a dynamic mathematical model which combines previously disconnected research traditions explicitly describing the coevolution of behavior and beliefs. Overall our results show that one can hardly understand behavior without understanding the dynamics of personal beliefs and beliefs about others and that cognitive, social, and material factors all play important roles in these processes.

In our experiment, the effects of all factors considered were statistically significant. The most important factors in decision-making were personal norms (cognitive dissonance) and empirical expectation (conformity with expected peers' actions), while the least important factors were expected material payoffs and normative expectations (expected disapproval by peers). Without messaging the weights of personal norms and conformity were $B_1 = 0.32$ and $B_3 = 0.16$. Messaging decreased the former to $B_1 = 0.19$, while increased the latter to $B_3 = 0.21$. It appears that authority's messaging relaxes the need to evaluate whether the action chosen fits the personal norm, while it simultaneous makes people to pay closer attention to how others react to it. The weights of material factors and of normative expectations were similar in the two experimental treatments: $B_0 = 0.11, B_2 = 0.09$ without messaging and $B_0 = 0.12, B_2 = 0.10$ with messaging. From previous experimental research, we know that people are often not exclusively motivated by their self-interests and can fail to maximize them. Nevertheless, that subjects did not do better in maximizing their material payoffs is especially surprising because in our experiment they had a tool for estimating their payoffs given their and others' actions and 35 days to learn using it. In real-life situations, people usually do not have such a "cheating gadget" which would make payoff maximization even more difficult. In our experiment, subjects lost up to between 50% and 15% of their payoffs due to non-material factors (see Fig. S21 in the SM). The low weight of normative expectations is likely due to the fact that subjects interacted in groups that were randomly formed at each round. We also showed that personal norms and beliefs about others can change rapidly (Tormos 2020). Personal norms equilibrated the fastest. In the dynamics of personal norms and normative expectations, cognitive factors were as important as conformity with observed peers' behavior. Although the weights we report here might be specific to our experiment, they provide a useful yardstick for future experimental studies using different games and subjects.

We did not observe any direct effect of messaging on the average actions. However, messaging increased variation between individuals: while pro-social types and rule-following types have reduced their extraction efforts, individualistic types and rule-breakers opportunistically increased their efforts enjoying higher payoffs. Individuals behaving cooperatively were characterized by higher weight of personal norms (B_1) while non-cooperative individuals had higher weight of conformity (B_3). Messaging was most important for personal norms (when its weight was three times that of the joint weight of cognitive dissonance and conformity), and for normative expectations (when its weight was comparable with that of the joint weight of cognitive factors and conformity), and the least important for empirical expectations. The messaging reduced the importance of cognitive factors (cognitive dissonance, social projections, logic constraints) and of conformity with peers on beliefs dynamics. As in the case of individual action, it appears that the presence of authority's messaging reduces the need to perform complex cognitive evaluations associated with changing beliefs. Overall however our results clearly show that cognitive factors are as important as material and social ones.

Our results support earlier conclusions about the importance of personal norms for social behavior (Basić and Verrina 2020, Capraro and Rand 2018, Catola et al. 2021, Schwartz 1977). In our subjects, the weight of personal norms is 50% to 100% larger than that of the second most important factor which is conformity. The deviations from personal norms are sanctioned via cognitive dissonance (Festinger 1957). We have also shown that personal norms can change as a result of social interactions and that for a substantial proportion of subjects they can be altered by messaging.

Individual differences were large across all characteristics studied (c.f. (Bicchieri and Dimant 2019, Sunstein 2021)). The distributions of most parameters were highly skewed with a number of extreme individuals. Similar to earlier work (Fehr and Schurtenberger 2018, Poncela-Casasnovas et al. 2016, Szekely et al. 2021), there were "behavioral types" different in the type of social information used. Antisocial individuals (i.e., those who invest more in the CPR extraction) are those who can be viewed as greedy (higher B_0), without principles and shame (low B_1), B_2, and conformists in behavior (high B_3). They also can have larger sensitivity to potential disapproval by others (B_2) than prosocial types who mostly act according to their personal norms. The gender differences were small.

In our model, the effects of material payoffs in decision-making were specified by the term $B_0\theta$ in equation (2), where θ is the action x maximizing payoff $\pi(x|\tilde{x})$ given the individual belief $\tilde{x}$ about the average expected action of group mates, and B_0 is the parameter estimated from data. Our subjects had a computer tool and a table to estimate θ, although of course we do not know how often they used it. The average estimate of B_0 was small. However for a number of subjects (11/27 in experiments without/with messaging, respectively) we could not estimate B_0 because they did not exhibit sufficient variation in θ. Some of these subjects made consistently high extraction effort so it is likely they were motivated by the payoffs. Moreover, our analysis has identified a number of subjects (such as "conditional coperators" and "conditional compliers", see the SM) for whom our model did not work well. All this means that our estimates of the weight of material factors B_0 may be biased downward. Future work using more complex models of actions may be able to resolve this issue.

Our conclusion about a low weight of material factors in decision-making may appear to be difficult to reconcile with the fact that average individual efforts converge to values just below $x = 22$ which are relatively close to the Nash equilibrium at $x_{\mathrm{NE}} = 24$. This however is a result of social influence (expressed via normative and empirical expectations) rather than of payoff maximization. A well-established fact in social influence theory is that the presence of even a single individual (or a motivated minority) who refuses to change their opinion can "pull" the whole group to match their opinion (Couzin et al. 2011, Flache et al. 2017, Gavrilets et al. 2016). Similarly in our case, a small number of people motivated by material payoffs "pull" the whole group closer to their own preference.

Our results validate some theoretical predictions made earlier in Gavrilets (2021). As predicted, without messaging $\tilde{x}$ appears to evolve to match x while $\tilde{y}$ appears to evolve to match y (Figs. S2 and S3). With messaging, the mean group effort evolves to a value between the Nash equilibrium and the value promoted by messaging $G = 14$. In both cases, the observed variances in the main dynamic variables are in the order predicted: $\sigma_x^2 > \sigma_y^2 > \sigma_{\tilde{y}}^2 > \sigma_{\tilde{x}}^2$. Also the lack of the effect of messaging on actions x can be viewed as a partial support of the "backfiring" identified by Gavrilets (2021) in the CPR game. However other predictions were not supported. For example, without messaging, the average extraction effort x did not appear to evolve to the Nash equilibrium $x_{\mathrm{NE}} = 24$, while personal norms y and normative expectations $\tilde{y}$ did not appear to evolve to match x. A possible reason for this mismatch is that our experiment was not long enough. An additional cause is the presence of subjects who did not change their personal norms during the curse of the experiment (i.e., subjects with $\alpha_1 = \beta_2 = 0$) as well as subjects whose decision-making was not captured well by our model (as discussed in the SM).

In our experiments, we explicitly asked subjects to think about their norms and beliefs before each choice. We also incentivized them when eliciting their normative and empirical expectations. We did not have a control experimental treatment without these prompts potentially making it difficult to assess the psychological importance of norms and beliefs in situations where such prompts are lacking but material benefits are present. This is a potential reason for concern. However, first, we think that the existing evidence on the effect of this "priming" is mixed. For example, in their review entitled "Belief elicitation in the laboratory", Schotter and Trevino (2014) conclude that "the process of eliciting beliefs seems not to be too intrusive" [p.103]. Schlag et al. (2015) discuss both disadvantages and advantages of incentivized methods for eliciting beliefs. D'Adda et al. (2016) who studied the effect of norm elicitation on behavior conclude "We also find little evidence that eliciting norms affects subsequent behavior in the bribery game" (p. 2). In our earlier paper on the collective risk game (Szekely et al. 2021), we did not observe a statistically significant difference between actions taken when beliefs were elicited before or after the action. In our paper (Rosokha et al. 2022), we explicitly looked at the effects of beliefs elicitation on behavior in a Public Goods Game and did not find them. Second, one can argue that in most real-life situations resembling social dilemmas, people are subject to different social influences. In other words, in real world people are primed to consider various social influences and also are hardwired to experience certain psychological forces pushing them to comply to a certain extent with the expectations of others.

Although the mathematical model underlying our approach is rather general, here we have tested it only using WEIRD (Western, Educated, and from Industrialized, Rich, and Democratic countries) subjects (Henrich 2020, Henrich et al. 2010, Muthukrishna et al. 2020) and only within the context of social behavior related to resource use (Ostrom et al. 1992, Walker et al. 1990). To confirm the generality of our finding it is necessary to study non-WEIRD populations as well as populations differing in tightness-looseness of their cultures (Chua et al. 2019, Gelfand et al. 2011, Harrington and Gelfand 2014, Jackson et al. 2020). It is also necessary to consider other types of social interactions such as coordination, production of public goods, actions to avoid collective risks, sharing of resources, rule-following etc. It would be particularly interesting to study games stimulating the emergence of stronger social norms (e.g., those implemented within contexts of high risk of catastrophic losses, (Szekely et al. 2021)); considering different types of messaging, especially those exploiting social identity and/or existing social norms; and studying real-life analogs and applications. It is important to generalize our theoretical approach by considering more complex and realistic models of learning (Chandrasekhar et al. 2020, Young 2004), more complex (e.g., asymmetric or discontinuous) utility functions as well as discrete rather than continuous action spaces (Bernheim 1994). Our analysis also highlights the need for more complex models of decision-making.

Pursuing these lines of research may lead to an integrative picture of human behavior, incorporating its coevolution with attitudes and beliefs about others in a manner closely related to actual social situations. Importantly, such picture would not provide a one-size-fits-all description of human behavior, but on the contrary, it will place the focus on between-individual differences and on the relevance of considering the diversity of people involved in a social process to understand it adequately.

From a practical perspective, our methods and results may suggest ways to better understand and predict behavior and beliefs change in a population as a result of policy interventions (e.g., aimed at environment protection or climate change mitigation) or certain shocks, such as an epidemic, a natural disaster, or social unrest (Bavel et al. 2020, Kuran 1995, Tankard and Paluck 2016, Thaler and Sunstein 2021). Our CPR experiment has quantified how descriptive, injunctive, and personal norms together with authority's messaging affect human behavior and how beliefs coevolve with actions. It also shows that different individuals respond differently to these forces. For example, a certain proportion of people accept authority's messaging while others do not. For some people, conformity with others is very important while others are mostly motivated by personal norms. From earlier work, we know that using multiple nudges simultaneously can increase the efficiency of messaging (Brandon et al. 2019). All this implies that policymaking must embrace heterogeneity and gather information about the distribution of individuals' reactivity to different (material, cognitive, social) factors in the population, in order to design and implement suitable intervention strategies for different contexts. One possibility can be to use messaging which targets different clusters of individuals (e.g., those who are driven mostly by personal norms or by conformity with others). Knowing the approximate proportions of different types in the whole population and their typical responses (which could be estimated in pilot experiments) might allow one to design more effective ways to allocate resources between different nudges. Another more involved option would be to develop highly personalized messaging (Milles 2020, Sunstein 2013). To realize these possibilities it is necessary to establish to what extent our approach and results are generalizable to other social situations and populations.

Data availability

The datasets and Matlab codes generated during and/or analyzed during the current study are available in the Zenodo repository https://zenodo.org/record/7853468#.ZEL40PdOlhE.

Received: 10 February 2023; Accepted: 2 May 2023;
Published online: 13 May 2023

References

Adeneye YB, Jaaffar AH, Ooi CA, Ooi SK (2021) Nexus between carbon emissions, energy consumption, urbanization and economic growth in Asia: evidence from common correlated effects mean group estimator (ccemg). Front Energy Res 8:610577

Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Processes 50:179–211

Akerlof GA, Dickens WT (1982) The economic consequences of cognitive dissonance. Am Econ Rev 72:307–319

Andreozzi L, Ploner M, Saral AS (2020) The stability of conditional cooperation: beliefs alone cannot explain the decline of cooperation in social dilemmas. Sci Rep 10:13610

Apesteguia J (2006) Does information matter in the commons? Experimental evidence. J Econ Behav Organ 60:55–69

Apesteguia J, Maier-Rigaud FP (2006) The tole of rivalry: public goods versus common-pool resources. J Conflict Resolut 50:646–663

Basić Z, Verrina E (2020) Personal norms—and not only social norms—shape economic behavior. Technical report, Max Planck Institute for Research on Collective Goods

Bavel J, Baicker K, Boggio P, Capraro V, Cichocka A, Cikara M, Crockett MJ (2020) Using social and behavioural science to support COVID-19 pandemic response. Nature Human Behavior 4:460–471

Belsley DA (1991) A guide to using the collinearity diagnostics. Comput Sci Econ Manag 4(1):33–50

Belsley DA, Kuh E, Welsch RE (2005) *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons

Benet-Martínez V, John OP (1998) Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the big five in spanish and english. J Personal Soc Psychol 75(3):729

Bernays E (1928) *Propaganda*. Ig Publishing

Bernheim B (1994) A theory of conformity. J Political Econ 102(5):841–877

Bicchieri C (2006) *The grammar of society. The nature and dynamics of social norms*. Cambridge University Press, Cambridge

Bicchieri C, Dimant E (2019) Nudging with care: the risks and benefits of social information. *Public Choice*, https://doi.org/10.1007/s11127-019-00684-6

Brandon A, List JA, Metcalfe RD, Price MK, Rundhammer F (2019) Testing for crowd out in social nudges: evidence from a natural field experiment in the market for electricity. Proc Natl Acad Sci 116(12):5293–5298

Burnham KP, Anderson DR (2001) Kullback-leibler information as a basis for strong inference in ecological studies. Wildl Res 28(2):111–119

Bursztyn L, Yang DY (2021) Misperceptions about others. Annu Rev Econ 14:425–452

Calabuig V, Olcina G, Panebianco F (2018) Culture and team production. J Econ Behav Organ 149:32–45

Capraro V, Rand DG (2018) Do the Right Thing: experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. Judgement Decis Making 13(1):99–111

Capraro V, Jordan J, Rand D (2014) Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. Sci Rep 4:6790

Catola M, D'Alessandro S, Guarnieri P, Pizziol V (2021) Personal norms in the online public good game. Econ Lett 207:10024

Centola D, Willer R, Macy M (2005) The emperor's dilemma: a computational model of self-enforcing norms. Am J Sociol 110:1009–1040

Chandrasekhar AG, Larreguy H, Xandri JP (2020) Testing models of social learning on networks: evidence from two experiments. Econometrica 88:1–32

Chua RYJ, Huang KG, Jin M (2019) Mapping cultural tightness and its links to innovation, urbanization, and happiness across 31 provinces in China. Proc Natl Acad Sci USA 116:6720–6725

Cialdini RB, Reno RR, Kallgren CA (1990) A focus theory of normative conduct:701recycling the concept of norms to reduce littering in public places. Personal Soc Psychol 58:1015–1026

Couzin ID, Krause J, Franks NR, Levin SA (2011) Effective leadership and decision-making in animal groups on the move. Nature 433:513–516

Croson R, Marks M (2001) The effect of recommended contributions in the voluntary provision of public goods. Econ Inquiry 39(2):238–249

D'Adda G, Drouvelis M, Nosenzo D (2016) Norm elicitation in within-subject designs: testing for order effects. J Behav Exp Econ 62:1–7

d'Adda G, Dufwenberg M, Passarelli F, Tabellin G (2020) Social norms with private values: theory and experiments. Games Econ Behav 124:288–304

Dal Bó E, Dal Bó P (2014) "Do the right thing:" the effects of moral suasion on cooperation. J Public Econ 117:28–38

Dave C, Eckel CC, Johnson CA, Rojas C (2010) Eliciting risk preferences: when is simple better? J Risk Uncertain 41(3):219–243

DeGroot M (1974) Reaching a consensus. J Am Stat Assoc 69:118–121

Dong K, Sun R, Hochman G (2017) Do natural gas and renewable energy consumption lead to less co2 emission? empirical evidence from a panel of brics countries. Energy 141:1466–1478

Fehr E, Schurtenberger I (2018) Normative foundations of human cooperation. Nat Hum Behav 2:458–468

Festinger L (1957) A theory of cognitive dissonance. Stanford University Press, Palo Alto, CA

Fischbacher U, Gächter S (2010) Social preferences, beliefs, and the dynamics of free riding in public goods experiments. Am Econ Rev 100:541–556

Fishbein M, Ajzen I (1975) *Belief, attitude, intention and behavior*. Addison-Wesley

Flache A, Mäs M, Feliciani T, Chattoe-Brown E, Deffuant G, Huet S, Lorenz J (2017) Models of social influence: towards the next frontiers. J Artif Soc Soc Simul 20(4):2

Friedkin NE, Proskurnikov AV, Tempo R, Parsegov SE (2016) Network science on belief system dynamics under logic constraints. Science 354:321–326

Fudenberg, D, Tirole J (1992) *Game theory*. The MIT Press, Cambride, MS

Galesic M, Stein DL (2019) Statistical physics models of belief dynamics: theory and empirical tests. Phys A: Stat Mech Appl 519:275–294

Galesic M, Olsson H, Dalege J, van der Does T, Stein DL (2020) Integrating social and cognitive aspects of belief dynamics: towards a unifying. J R Soc Interface 18:2020085

Gavrilets S (2020) The dynamics of injunctive social norms. Evoluti Hum Sci 2:e60

Gavrilets S (2021) Coevolution of actions, personal norms, and beliefs about others in1social dilemmas. Evolut Hum Sci 3:e44

Gavrilets S, Auerbach J, van Vugt M (2016) Convergence to consensus in heterogeneous groups and the emergence of informal leadership. Sci Rep 6:29704

Gavrilets YN (2003) Stochastic modeling of between-group social interactions. Econ Math Methods 39:106–116

Gelfand MJ et al. (2011) Differences between tight and loose cultures: a 33-nation study. Science 332:1100–1104

Górges L, Nosenzo D (2020) Measuring social norms in economics: why it is important and how it is done. Analyse Kritik 42:285–311

Guido A, Tverskoi D, Gavrilets S, Sánchez A, Andrighetto G (2023) Nudging or nagging: The perils of persuasion. Available at SSRN: https://ssrn.com/abstract=4404960 or https://doi.org/10.2139/ssrn.4404960

Harrington JR, Gelfand MJ (2014) Tightness-looseness across the 50 united states. Proc Natl Acad Sci USA 111:7990–7995

Henrich J (2020) The WEIRDest people in the world. How the West became psychologically peculiar and particularly prospective. Farrar, Straus, and Giroux, New York

Henrich J, Heine S, Norenzayan A (2010) Most people are not WEIRD. Nature 466:29

Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for non-orthogonal problems. Technometrics 12(1):55–67

Jackson JC, Gelfand M, Ember CR (2020) A global analysis of cultural tightness in non-industrial societies. Proc R Soc Lond B 287:20201036

John OP, Donahue E, Kentle R (1991) The big five inventory: Versions 4a and 54 [technical report]. *Berkeley: University of California, Institute of Personality and Social Research*

Kashima Y, Laham SM, Dix J, Levis B, Wong D, Wheeler M (2015) Social transmission of cultural practices and implicit attitudes. Organ Behav Hum Decis Process 127:113–125

Kashima Y, Perfors A, Ferdinand V, Pattenden E (2021) Ideology, communication and polarization. Philos Trans R Soc Lond B 376:20200133

Kimbrough EO, Vostroknutov A (2016) Norms make preferences social. J Eur Econ Assoc 14:608–638

Kimbrough EO, Vostroknutov A (2018) A portable method of eliciting respect for social norms. Econ Lett 168:147–150

Kölle F, Quercia S (2021) The influence of empirical and normative expectations on cooperation. J Econ Behav Organ 190:691–703

Krueger JI (2007) From social projection to social behaviour. Eur Rev Soc Psychol 18:1–35

Kuran T (1995) *Private truths, public lies*. Harvard University Press, Cambridge

Kuran T, Sandholm WH (2008) Cultural integration and its discontents. Rev Econ Stud 75(1):201–228

Lee Y, Sul D (2022) Trimmed mean group estimation. In Chudik A, Hsiao C, and Timmermann A (eds) *Essays in honor of M. Hashem Pesaran: panel modeling, micro applications, and econometric methodology*, Emerald Publishing Limited, p 177–202

Loewenstein G, Molnar A (2018) The renaissance of belief-based utility in economics. Nat Human Behav 2:166–167

Marlowe FW (2004) What explains Hadza food sharing? Res Econ Anthropol 23:69–88

Marx K (1959) A contribution to the critique of political economy. Charles H Kerr and Company, Chicago

Mertens S, Herberz M, Hahnel UJJ, Brosch T (2022) The effectiveness of nudging: A meta-analysis of choice architecture interventions across behavioral domains. Proc Natl Acad Sci 119(1):e2107346118

Milles S (2020) Personalized nudging. *Behavioural Public Policy*, https://doi.org/10.1017/bpp.2020.7

Molnar A, Loewenstein G (2022) Thoughts and players: an introduction to old and new economic perspectives on beliefs. In "The Cognitive Science of Belief: A Multidisciplinary Approach" edited by Julien Musolino, Joseph Sommer, Pernille Hemmer. pp. 21–350. Cambridge University Press

Murphy RO, Ackermann KA (2014) Social value orientation theoretical and measurement issues in the study of social preferences. Personal Soc Psychol Rev 18:13–41

Murphy RO, Ackerman KA, Handgraaf MJJ (2011) Measuring social value orientation. Judgment Decis Making 6:771–781

Muthukrishna M, Bell AV, Henrich J, Curtin CM, Gedranovich A, McInerney J, Thue B (2020) Beyond western, educated, industrial, rich, and democratic (WEIRD) psychology: measuring and mapping scales of cultural and psychological distance. Psychol Sci 31:678–701

Ndambendia H, Njoupouognigni M (2010) Foreign aid, foreign direct investment and economic growth in sub-saharan africa: evidence from pooled mean group estimator (pmg). Int J Econ Finan 2(3):39–45

Olsson A, Knapska E, Lindström B (2020) The neural and computational systems of social learning. Nat Rev Neurosci 21:197–212

Ostrom E, Walker J, Gardner R (1992) Covenants with and without a sword: self-governance is possible. Am Political Sci Rev 86:404–417

Paramati SR, Roca E (2019) Does tourism drive house prices in the OECD economies? Evidence from augmented mean group estimator. Tour Manag 74:392–395

Pesaran H, Smith R (1995) Estimating long-run relationships from dynamic heterogeneous panels. J Econ 68:79–113

Piotrowski EW, Sladkowski J (2003) An invitation to quantum game theory. Int J Theor Phys 42:1089–1099

Poncela-Casasnovas J, Gutierrez-Roig M, Gracia-Lazaro C, Vicens J, Gomez-Gardenes J, Perello J, Moreno Y, Duch J, Sánchez A (2016) Humans display a reduced set of consistent behavioral phenotypes in dyadic games. Sci Adv 2:1600451

Premack D, Woodruff G (1979) Does the chimpanzee have a theory of mind. Behav Brain Sci 1:515–526

Rabin M (1994) Cognitive dissonance and social change. J Econ Behav Organ 24:177–194

Rashevsky N (1949) Mathematical biology of social behavior. III. Bull Math Biol 11:255–271

Rawlings CM (2020) Cognitive authority and the constraint of attitude change in groups. Am Sociol Rev 85:992–1021

Redner S (2019) Reality inspired voter models: a mini-review. Comptes Rendus Physique 20:275–292

Rosokha Y, Lyu X, Tverskoi D, Gavrilets S (2022) Evolution of cooperation in the indefinitely repeated collective action with a contest for power. Available at SSRN: https://ssrn.com/abstract=4206122 or https://doi.org/10.2139/ssrn.4206122

Sadorsky P (2013) Do urbanization and industrialization affect energy intensity in developing countries? Energy Econ 37:52–59

Sandholm WH (2010) *Population games and evolutionary dynamics*. MIT Press, Cambridge, Massachusetts

Schlag KH, Tremewan J, van der Weele JJ (2015) A penny for your thoughts: a survey of methods for eliciting beliefs. Exp Econ 18:457–490

Schotter A, Trevino I (2014) Belief elicitation in the laboratory. Annu Rev Econ 6:103–128

Schwartz SH (1977) Normative influences on altruism. Adv Exp Soc Psychol 10:221–279

Sunstein C (2013) Behavioral economics and paternalism. Yale Law J 122:1867–1899

Sunstein CR (2021) The distributional effects of nudges. *Nat Hum Behav*. https://doi.org/10.1038/s41562-021-01236-z

Szekely A, Lipari F, Antonioni A, Paolucci M, Sánchez A, Tummolini L, Andrighetto G (2021) Collective risks change social norms and promote cooperation: Evidence from a long-term experiment. Nat Commun 12:5452

Tankard ME, Paluck EL (2016) Norm perception as a vehicle for social change. Soc Issues Policy Rev 10:181–211

Teal F, Eberhardt M (2010) Productivity analysis in global manufacturing production. Working paper 515. Department of Economics Discussion Paper Series. University of Oxford. https://ora.ox.ac.uk/objects/uuid:ea831625-9014-40ec-abc5-516ecfbd2118

Tembine H (2017) Mean-field-type games. AIMS Math 2:706–735

Thaler RH, Sunstein CR (2021) *Nudge: the final edition*. Penguin

Tormos R (2020) *The rhythm of modernization. how values change over time*. Brill, Leiden

Walker JM, Gardner R, Ostrom E (1990) Rent dissipation in a limited-access common-pool resource: experimental evidence. J Environ Econ Manag 19:203–211

Watts DJ (2002) A simple model of global cascades on random networks. Proc Natl Acad Sci USA 99:5766–5771

Young HP (2015) The evolution of social norms. Annu Rev Econ 7:359–387

Young P (2004) *Strategic learning and its limits*. Oxford University Press

Acknowledgements

We thank A. Cabrales, L. Gaertner, Y. Rosokha, and L. Tummolini and reviewers for comments and suggestions and P. Lozano for help with programming. GA was supported by the Knut and Alice Wallenberg Foundation Grant 2016.0167 and the Research Project of National Relevance (PRIN) "14ALL", funded by the Italian Ministry of Education, University and Research. AS was supported by grants PGC2018-098186-B-I00 (BASIC, FEDER/MICINN- AEI) funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe" and PRACTICO-CM (Comunidad de Madrid). SG was supported by the U. S. Army Research Office grants W911NF-14-1-0637 and W911NF-18-1-0138, the Office of Naval Research grant W911NF-17-1-0150, the Air Force Office of Scientific Research grant FA9550-21-1-0217, and the John Templeton Foundation.

Competing interests

The authors declare no competing interests.

Ethical approval

The study complied with all relevant ethical regulations for work with human participants. The study received institutional ethical approval from the Ethics Committee of the Universidad Carlos III de Madrid.

Informed consent

Informed consent was obtained from all subjects.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-023-01745-4.

Correspondence and requests for materials should be addressed to Sergey Gavrilets.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.