Do Norm-Nudges Induce Excessive Pro-Social Behavior and Severe Punishment?: An Experimental Study on Asymmetric Public Goods Games [*]

Tetsuo Yamamori[a] Tadakatsu Nakamura[b]

Abstract

We explore whether norm-nudges induce excessive pro-social behavior and severe punishment by conducting a laboratory experiment based on an asymmetric public goods game with punishment. We find that a moral message induces both excessive contribution (those over the social optimal level) and severe punishment, whereas a social comparison nudge does not induce excessive contribution, but induces severe punishment. Consequently, both nudge messages worsen social welfare.

Keywords: norm-nudge, social norm, asymmetric public goods game, punishment, economic experiment

JEL classification: C92, D90, H41

[*] This research was financially supported by JSPS KAKENHI [Grant Number 23K01402]. We have no conflicts of interest to disclose.

[a] Faculty of Economics, Dokkyo University, 1-1, Gakuen-cho, Soka, Saitama 340-0042, Japan, yamamori@dokkyo.ac.jp.

[b] Faculty of Regional Policy, Takasaki City University of Economics, 1300 Kaminamie, Takasaki, Gunma, 370-0801, Japan, tadakatu@tcue.ac.jp.

1. Introduction

Nudges to promote pro-social behavior in situations involving conflict between individual interests and public welfare, such as social dilemmas, relying on the assumption that people obey social norms are called “norm-nudges” (Bicchieri and Dimant 2022). This is typically done by issuing messages about what most people in the same situation do and/or what most people in the same situation approve of, or directly stating what the right thing to do in this situation is. These nudges attempt to manipulate people’s social expectations to shift existing social norms or to create new norms.

In this study, based on laboratory experiments, we explore the possibility that norm-nudges worsen social welfare by inducing people for whom the costs associated with complying with social norms are high to engage in excessive pro-social behavior or inducing others to punish such people harshly. Our starting point is the fact that there are vast differences among people in the marginal costs incurred by an individual in complying with a social norm more often or intensely, and in the marginal benefits that an individual derives from another person’s conformity to the norm. Furthermore, since social norms are supported by the threat of informal sanctions (Elster 1989; Fehr and Fischbacher 2004), norm-nudges may influence our expectations about what kind of behavior will elicit others’ punishments and their severity. Therefore, as a result of a norm-nudge, a certain type of behavior becomes strongly recognized as a social norm, and some people probably commit such behavior without considering their marginal costs out of fear of severe punishment by others. In other words, norm-nudges may induce some people to engage in excessive pro-social behavior in social dilemmas.

Whether a norm-nudge induce some people to engage in excessive pro-social behavior, and what consequences are for social welfare are empirical questions. To date, considerable field research has found that some messages based on norm-nudges promote pro-social behaviors in social dilemmas, involving saving electricity, conserving water, and taking preventive actions against the spread of an infection. (e.g., Allcott 2011; Ferraro and Price 2013; Sasaki et al. 2021). However, through field research, it is difficult to identify people’s marginal costs/benefits associated with such behaviors, and so to social optima and define what degrees of commitments to pro-social behavior are excessive. Therefore, laboratory experiments are conducted to address these questions.

2. Experimental design and procedure

Our experiments are based on a version of McGinty and Milam’s (2013) asymmetric public goods game. There are n players, each of which has the same amount of endowment e. Each player i must divide their endowment between a contribution y_i to the joint project (public goods) shared with other players and investment $e - y_i$ in their own business, which solely generates private benefits. The payoffs of each player i earn from their joint project are determined by the following function, which has diminishing marginal returns from contributions.

$$G_i(Y) = b\alpha_i\left(aY - \frac{Y^2}{2}\right). \tag{1}$$

where a and b are non-negative scale parameters, $Y = \sum_{i=1}^{n} y_i$ is the aggregate contribution across all players, and $\alpha_i > 0$ is the individual benefit parameter. The payoffs of each player earning from their investment $e - y_i$ for their own business are determined by the following function, which implies an increasing marginal opportunity cost from public contributions y_i.

$$P_i(y_i) = c_i\left(d(e - y_i) - \frac{(e - y_i)^2}{2}\right) \tag{2}$$

where d is a non-negative scale parameter, and $c_i > 0$ is the individual parameter that allows for asymmetry in players' opportunity costs from public contributions. Y_{-i} is denoted as $\sum_{j \neq i} y_j$. Then, total payoffs of player i is as follows:

$$u_i(y_i, Y_{-i}) = G_i(Y) + P_i(y_i). \tag{3}$$

The underlying game in our experiments is a version of this game with two decision stages: choosing the contribution amounts to public goods (Stage 1) and deciding who to punish and how much to punish (Stage 2). There are three players ($n = 3$), one of whom has high marginal costs to contributions and low benefits from public goods (H player), and the rest have low marginal costs to contributions and high benefits from public goods (L player). Table 1 presents our experimental parameters, where y_i^* and y_i^o are player i's level of contribution in the Nash equilibrium and their social optimal level of contribution that maximizes the total payoffs, respectively. We define the parameters such that y_i^* and y_i^o exist uniquely and are both strictly interior to the endowment space.

Table 1. Experimental parameters

Types of players	# of each type	e	d	a	b	α_i	c_i	y_i^*	y_i^o
H: high cost, low benefit	1	8	8	24	20	0.2	60	1	3
L: low cost, high benefit	2					0.4	30	4	6

In Stage 1, each player chooses a contribution $y_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ to their joint project, knowing their own type and the others' type(s) but without knowing others' contributions (simultaneous move). By restricting y_i to an integer, all players payoff *points* (experimental currency unit) determined by equation (3) are integers.

In Stage 2, each player receives an additional 400 points and can punish the other two players using these points, knowing their contributions in the first stage. Let P_j^i is the punishment points that player i uses for punishing player $j \neq i$. Then $3P_j^i$ are reduced from the payoffs that player

j earned in Stage 1. Therefore, overall payoffs of player i in this game are given by

$$u_i(y_i, Y_{-i}) - 3\sum\nolimits_{j \neq i} P_i^j - \left(400 - \sum\nolimits_{j \neq i} P_j^i\right). \tag{4}$$

Our experimental design consisted of three treatments: the baseline public goods game (PG) and two nudge treatments: moral message treatment (MM) and social comparison treatment (SC). All treatments were essentially the same, except that a message based on norm-nudges (hereafter, nudge message) was written in the instructions and displayed on the computer screen of each nudge treatment. The messages in the MM and SC are as follows:

(MM) You should consider not only your profits but also the profits of your group members and try to contribute a significant number of tokens into the joint project for them.

(SC) Of past participants in this experiment, the largest number contributed X tokens to their joint account.

The message in the MM is a type of norm-nudge telling people the right thing to do with altruistic messages, whereas the message in the SC is a type of norm-nudge telling people about what most people in the same situation do, which is also called a social comparison nudge. Based on the experimental results of PG sessions conducted up to the first SC session, we set X in the SC message to four.

Our experiments were conducted at the Dokkyo University and the Takasaki City University of Economics, Japan, between November 2023 and July 2024. We recruited participants by displaying posters and distributing fliers. The participants were undergraduate students from several departments who had not participated in any prior public goods experiments. Each participant could take part in only one session. The total number of participants for the PG, MM, and SC was 93, 114, and 66, respectively.

Each session was conducted in a computer room, and the z-tree software package developed by Fischbacher (2007) was used. The participants were randomly seated in front of a computer terminal. Each desk was surrounded by partitions and contained all the experimental materials, including instructions, practice problems, and an ID card. IDs were randomly assigned to each desk in advance. The roles of each participant (H player or L player) and the members of the same group were determined using their IDs. To avoid potential experimenter effects, assistants other than researchers acted as instructors. The instructors read instructions aloud. Before the experiment commenced, the participants were instructed to solve practice problems to verify their understanding of the experiment instructions. The experiment began only after all participants provided correct answers to practice problems. Five rounds were conducted in each session to allow for learning effects. The role of each participant remained fixed during the session; however, group members were regrouped in each round, and the probability of meeting the same participants again as the same group member was zero. All decisions were anonymous, and the participants did not know the personal identities of their group members.

3. Results

Table 2 shows the means of the experimental outcomes by treatment in Stage 1. The nudge message of the MM seems to promote contributions to public goods (not unduly); the means of the contributions of H and L players in the MM do not exceed y_i^o but are significantly larger than those in the PG (Mann–Whitney U test, $p < 0.01$ for each). Consequently, the total profit (the sum of the three players' profits) in Stage 1 of the MM is significantly larger than that of the PG (Mann–Whitney U test, $p < 0.01$). However, in the MM, the proportion of H players who contributed more than y_i^o is double that in the PG: the means of the excessive contribution dummy (equals 1 if $y_i > y_i^o$, 0 otherwise) in the MM is 0.23, while that in the PG is 0.11. This difference is statistically significant (Fisher's exact test, $p < 0.01$). A similar trend exists for L players, but its value is only 0.07 even in the MM. Therefore, although the nudge message in the MM increased the profits of L players in Stage 1, it had no effect on increasing the profits of H players. However, the nudge message of the SC seemed not to be effective in promoting contributions to public goods: the mean of the contributions of H players in the SC is larger than that in the PG, but the difference is not significant. For L players, the mean of the contributions in the SC is less than that in the PG. Consequently, the total profit of the SC in Stage 1 is slightly lower than that of the PG.

Table 2. Mean comparisons of experimental outcomes in Stage 1

		PG		MM		SC	
Contributions	H	1.76	(1.17)	2.54	(1.25)	2.10	(1.11)
	L	4.04	(0.98)	4.39	(1.30)	3.87	(1.16)
Excessive contribution	H	0.11	(0.31)	0.23	(0.42)	0.14	(0.34)
dummy	L	0.02	(0.15)	0.07	(0.26)	0.02	(0.15)
Profit in Stage 1	H	2530.35	(126.94)	2499.48	(180.93)	2494.96	(144.10)
	L	2189.53	(179.84)	2283.69	(193.11)	2202.99	(188.75)
Total profit in Stage 1		6909.39	(284.96)	7066.87	(272.07)	6900.96	(336.50)

Note: Values in parentheses are standard deviations.

Table 3 shows the means of the punishment points (the participants used to reduce others' payoffs earned in Stage 1) and overall profits. Despite the contributions to public goods being greater in MM than in PG, the punishment points used by each of the H and L players in the MM were significantly higher than those in PG (Mann–Whitney U test, $p < 0.01$ for each). The nudge message of the SC had a similar effect on the participants' punishment behavior as that of the MM, although it had no effect on their contributions. Consequently, the final total profit (the sum of the three players' overall profits) is highest for the PG, followed by the MM and SC in that order. The differences in them between the PG and MM and between the MM and SC are both significant (t-test, $p < 0.01$ and $p < 0.1$, respectively). Therefore, we can conclude that, in our public goods game, nudge messages, whether moral or social comparison messages, worsen social welfare.

Table 3. Mean comparisons of punishment points and final profits

		PG		MM		SC	
Punishment points	H	14.36	(59.93)	45.96	(103.07)	33.64	(79.07)
	L	30.27	(67.74)	59.64	(109.11)	70.42	(121.49)
Overall profit	H	2780.57	(235.35)	2602.50	(359.36)	2597.71	(369.53)
	L	2514.60	(226.74)	2501.70	(318.19)	2402.66	(365.04)
Final total profit		7809.77	(512.83)	7605.90	(846.24)	7403.03	(918.87)

Note: Values in parentheses are standard deviations.

References

Allcott, H. 2011. Social norms and energy conservation. Journal of Public Economics 95(910), 1082-1095.

Bicchieri, C. and E. Dimant, 2022. Nudging with care: the risks and benefits of social information. Public Choice 191(3), 443-464.

Elster, J. 1989. Social norm and economic theory. Journal of economic perspective 3(4), 99-117.

Fehr, E. and U. Fischbacher, 2004. Third-party punishment and social norms. Evolution and Human Behavior 25(2), 63-87.

Ferraro, P. J. and M. K. Price, 2013. Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. The Review of Economics and Statistics 95(1), 64-73.

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

McGinty, M. and G. Milam, 2013. Public goods provision by asymmetric agents: experimental evidence. Social Choice and Welfare 40(4), 1159-1177.

Sasaki, S., H. Kurokawa., F, Ohtake, 2021. Effective but fragile? Responses to repeated nudge-based messages for preventing the spread of COVID-19 infection. The Japanese Economic Review 72, 371-408.