ソシオネットワーク戦略ディスカッションペーパーシリーズ ISSN 2434-9445
第 109 号　2024 年 2 月
RISS Discussion Paper Series
No.109　February, 2024

Self-Regulatory Resources and Institutional Formation: An Experiment

Kenju Kamei

文部科学大臣認定　共同利用・共同研究拠点

関西大学ソシオネットワーク戦略研究機構

Research Institute for Socionetwork Strategies,
Kansai University
Joint Usage / Research Center, MEXT, Japan
Suita, Osaka, 564-8680, Japan
URL: https://www.kansai-u.ac.jp/riss/index.html
e-mail: riss@ml.kandai.jp
tel. 06-6368-1228
fax. 06-6330-3304

Self-Regulatory Resources and Institutional Formation: An Experiment

Kenju Kamei

Faculty of Economics, Keio University, 2-15-45, Mita, Minato-ku, Tokyo 108-8345, Japan

RIETI, 1-3-1, Kasumigaseki, Chiyoda-ku, Tokyo 100-8901, Japan

Email: kenju.kamei@gmail.com; kenju.kamei@keio.jp

Abstract: A novel laboratory experiment is used to show that the state of people's self-regulatory resources influences their reliance on the formal enforcement of norms in a social dilemma. The subjects' self-regulatory resources are manipulated using well-known depletion tasks. On the one hand, when their resources are not depleted, most decide to govern themselves through decentralized, peer-to-peer punishment in a public goods dilemma, and then achieve high cooperation norms. On the other hand, when the resources are limited, the majority enact a costly formal sanctioning institution; backed by formal punishment, the groups achieve strong cooperation. A supplementary survey on the Covid-19 pandemic was conducted to enhance the external validity of the findings, generating a similar pattern while revealing that people's desire to commit, not their beliefs about others' behavior without formal enforcement, drives their institutional preferences. Self-control preference theories, combined with inequity aversion, can explain these patterns because they predict that those with limited self-control are motivated to remove temptations in advance as a commitment device.

JEL codes: C92, D02, D72, D91, H41

Keywords: Institutional Choices, Social Dilemma, Public Goods, Self-Control, Punishment

Acknowledgement: This study was conducted as part of the project, "Self-regulatory resources and collective institutional choices in a social dilemma (an experimental study)," undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study was financially supported by a grant-in-aid from the Foundation for the Fusion of Science and Technology, and JSPS KAKENHI Grant Numbers 23H00833 and 23H00803. Keio University provided additional funding. The author thanks the members in the RISS at Kansai University, especially Kazuhito Ogawa, Toshihiko Nanba, Atsuko Shimano, and Hiroe Oshima, for their support in recruiting subjects, preparing for the experiment sessions, and managing the computers and software system in the laboratory; Artem Nesterov, for his assistance in zTree programming; and Xin Fang, for translating zTree program files from English into Japanese. The author is also grateful for the helpful comments and suggestions by Louis Putterman and the participants in the seminars at RIETI, Keio University, Kochi Institute of Technology, Chuo University, Waseda University.

1. Introduction

Human societies and organizations experience conflicts between private interests and socially optimal behaviors. Free-riding problems in social dilemmas are typical examples of such conflicts. In a social dilemma, people may recognize the value of cooperation and wish to achieve the Pareto-efficient outcome(s) through mutual cooperation. However, the temptation to free ride may be too strong for some to resist because of their self-control capacities (e.g., Gul and Pesendorfer, 2001, 2004; Baumeister *et al.*, 1994, 2007). Societies can regulate opportunistic behavior by implementing formal institutions (e.g., Ostrom, 1990), thus removing harmful temptations as a commitment device in advance. However, it is unclear how people's self-control capacities are linked to institutional formation in their community, whether in groups, societies, or organizations.

Over the past few decades, experimental studies have actively examined how formal (*a.k.a.* centralized) institutions can resolve social dilemmas and when these institutions should be implemented (e.g., Falkinger *et al.*, 2000; Tyran and Feld, 2006; Kosfeld *et al.*, 2009; Putterman *et al.*, 2011; Traulsen *et al.*, 2012; Zhang *et al.*, 2014; Kamei *et at.*, 2015; Nicklisch *et al.*, 2016; Fehr and Williams, 2018; Kamei and Tabero, 2021). Prior research suggests that not only do formal sanctioning institutions theoretically alter people's materially beneficial behaviors, but they also induce real people to make socially optimal choices (e.g., Falkinger, 1996; Falkinger *et al.*, 2000; Putterman *et al.*, 2011). However, formal institutions may not always be required to resolve dilemmas because people may successfully govern themselves through decentralized monitoring and peer-to-peer punishment (e.g., Fehr and Gächter, 2000, 2002; Masclet *et al.*, 2003; Gürerk *et al.*, 2006; Herrmann *et al.*, 2008; Gächter *et al.*, 2008; Casari and Luini, 2009; Ertan *et al.*, 2009). Several studies have investigated people's choices between formal and informal sanctioning institutions and found that groups prefer to use a formal institution over a decentralized solution under certain conditions, such as when the use of a formal institution does not entail a high cost (e.g., Kamei *et at.*, 2015), when anti-social peer-to-peer punishment is more severe than possible enforcement errors by a centralized authority (e.g., Nicklisch *et al.*, 2016), or when a normative consensus is difficult to reach through a decentralized mechanism (e.g., Fehr and Williams, 2018). This paper proposes and experimentally demonstrates another condition for formal institutions in social dilemmas: people's self-control capacity space or, more precisely, the amount of available *self-regulatory resources*.

Self-regulatory resources are internal energy that people use to regulate their behavior, thoughts, and emotions, and are commonly used for various kinds of self-regulation (e.g., resisting temptation and impulse, controlling emotions and thoughts, coping with stress and attention, and

dealing with conflicts between selfish and pro-social motivations). A large number of experiments in neighboring fields of economics have consistently demonstrated since around 1990 that (a) the state of their self-regulatory resources strongly influences people's decision-making, and (b) self-regulatory resources are *limited*, meaning that the resources are depleted once used for some activities (see, e.g., Baumeister *et al.* [1994, 2007] and Muraven and Baumeister [2000] for a survey). A typical psychology experiment uses a "sequential task paradigm" in which subjects first engage in a task with effortful control, followed by a different, unrelated task that requires self-control (e.g., Vohs and Baumeister, 2016). The first task aims to deplete subjects' self-regulatory resources, and there are various kinds of tasks (e.g., the Stroop (1935) color-naming task, a task to cross out letters under certain constraints, suppressing thoughts or emotions when watching a film, a task to forcing them to control attention, a task to force self-control).[1] Much more than 200 experiments have demonstrated that active regulation leads to self-control failure in later tasks (Vohs and Baumeister, 2016). Responding to the solid empirical evidence and great economic importance of self-control (e.g., overeating, consumption, borrowing, and procrastination), economists have joined the research and mathematically formalized people's self-control preferences (e.g., Gul and Pesendorfer, 2001, 2004; Fudenberg and Tirole, 2006; Dekel *et al.*, 2009). Recently, economic experiments have also verified that self-control preferences are indeed prevalent and that some people may want to remove strong temptations in advance if they anticipate that they will succumb to them and if removing them may improve their welfare (e.g., Bucciola *et al.*, 2011; Burger *et al.*, 2011; Houser *et al.*, 2018; Toussaert, 2018; Kocher *et al.*, 2017). Such self-control and commitment theories may also apply to the context of institutional formation in societies and organizations.

How to implement a formal institution is a difficult but important issue. Although this question is ubiquitous and commonly raised in modern societies, identifying people's institutional preferences and the effects of policies is challenging. One example is the restrictions related to the Covid-19 pandemic (which began in early 2020). Several countries, such as those in North America, Europe and Asia, enacted lockdowns or similar restrictions when the pandemic became serious. People's behavioral patterns during the pandemic resembled a self-regulatory depletion phenomenon. For example, Japan declared a state of emergency four times in Tokyo, and implemented strict restriction measures.[2] However, based on the data on people's flow, the impact of such restrictions diminished over time. For instance, on the first weekends following the declaration of the second,

[1] Self-regulatory depletion is related to a change in blood glucose. For example, Gailliot and Baumeister (2007) and Gailliot *et al.* (2007) showed that self-regulation acts cause a reduction in the blood-glucose levels.
[2] The first state of emergency was from April 7 to May 25, 2020; the second one was from January 8 to March 21, 2021; the third was from April 25 to June 20, 2021; and the fourth was from July 12 to Sep. 30, 2021.

third, and fourth states of emergency, the crowd numbers in Shibuya Center Street were found to be 50, 50, and 88 percentage points larger, respectively, than those on the first weekend following the declaration of the first state of emergency (Rei Frontier, Inc., July 2021). The news repeatedly announced that this kind of phenomenon was due to "Jishuku zukare" (exhaustion from extreme self-control, e.g., staying home) and emphasized that the state of emergency had become increasingly less effective. A survey conducted by the Cabinet Office in Spring 2021 indicated that 71.6% of the respondents agreed that they were exhausted from self-control.[3] Something similar occurred in almost every country. In the United Kingdom (UK), many citizens strictly followed social distancing measures and wore face coverings during the first lockdown. However, they gradually stopped following these measures or government recommendations. They even tended to oppose the restrictions when another wave later came. A survey, for example, showed that the percentage of those willing to self-isolate, if advised, decreased from 95% in April 2020 to 87% in April 2021 (Imperial College London, 2021). Parallel to the people's attitudes, the country gradually shifted toward living with the coronavirus without (strong) restrictions.

Nonetheless, this kind of interpretation may be misleading, because the Covid-19 restriction measures were weaker in later lockdowns/states of emergency. Thus, the pattern described above may simply mean that people's self-control is positively correlated with the strength of restriction measures. This opposite causation is similar to the well-known example of endogeneity demonstrated by Levitt (1997) for the positive correlations between crime rates and police force sizes in cities in the United States. As policymaking in democratic countries such as Japan and the UK reflects people's views, the weaker restriction measures in a later Covid-19 wave may mean that, *contrary* to self-control theory, people do not have commitment preferences when their self-regulatory resources are limited (i.e., when they cannot resist the temptation to go out due to, perhaps, self-regulatory depletion). Identifying people's commitment preferences is complex, nevertheless. For example, some unobserved individual characteristics, or omitted variables, might affect both people's self-control behaviors and their support for weak restriction measures through democratic processes. Uncertainty about the fatality of the coronavirus was also gradually resolved over time, which made comparisons of revealed behaviors between different time points less straightforward. People's concerns were not limited to their health and safety, as Covid-19 restrictions also impacted labor markets and incomes. Indeed, there is some indication of people's commitment preferences: Conducted in June 2021, an opinion survey by the Yomiuri newspaper found that (a) the percentage of those who supported changing the Japanese constitution increased from 49% in 2020 to 56% in

[3] https://www5.cao.go.jp/keizai2/wellbeing/covid/pdf/result3_covid.pdf (in Japanese; accessed on Nov. 5, 2022)

2021, and (b) 59% of the respondents agreed that the government's control rights and power should be strengthened in an emergency such as the Covid-19 crisis. In June 2021, an opinion survey by Jiji Press indicated that 53.7% (20.7%) of their respondents agreed (disagreed) to creating a clause in the constitution to strengthen the government's power in an emergency. However, whether individuals with strong or weak self-control support stronger formal enforcement is unclear. In addition, "emergency" in these surveys includes not only the Covid-19 crisis, but also any other crisis, such as a possible war with a neighboring country or a natural disaster.[4]

In addition to countermeasures for infectious diseases, there are many real-world situations in which institutional formation may depend on members' self-regulatory resources, whether in organizations such as firms or societies. For instance, each country faces environmental problems ranging from global (e.g., climate change, marine pollution, plastic waste, and biodiversity crisis) to local (e.g., water pollution and littering) problems. They also face other problems involving cooperation, such as trading, export control, protecting intellectual property, preventing cybersecurity, and promoting commerce. Conditions under which formal enforcement to restrict business activities within countries or across borders (e.g., enacting laws, strengthening environmental standards, enforcing fines for non-compliance) is desirable may depend on the amount of self-regulatory resources of those involved, because informal enforcement may suffice to regulate activities if self-governance and voluntary restraints are possible. One may also wonder whether the regulatory authority should limit firms' business activities in their customer relations, for example, in their sales or financing activities, to prevent them from exploiting customers. Residents in almost every community have some dilemmas (e.g., littering, noise) to their living environments, and it is sometimes observed that laws, referendums, or rules are introduced. However, when should people rely on formal enforcement rather than decentralized approach, such as monitoring? The present

[4] A similar difficulty arises when this research question is examined based on an existing cross-country dataset. For example, the World Value Survey Wave 7 (2017-2020) collected responses regarding what children were encouraged to learn at home, such as good manners and tolerance (Q7 to Q17 of the survey). As people are known to build self-regulatory resources in their lifetime (e.g., Baumeister *et al.*, 1994, 2007; Muraven and Baumeister, 2000), education to exercise self-control in early stages can be a proxy for their self-regulatory strength as a nation. The survey also collected views on government interventions by asking respondents to rate them on a 10-point scale: 1 = The government should take more responsibility to ensure that everyone is provided for, and 10 = People should take more responsibility to provide for themselves. A pairwise Pearson's ρ between the percentage of affirmative answers in Q7 to Q17 and the view on the government intervention was calculated as significantly negative ($\rho = -0.0247$, $p < 0.0001$). Thus, those more educated to build self-control in their childhood appear to ultimately support greater government responsibility. Furthermore, those who are more educated in self-control are significantly less confident with the current level of law in their nations according to the data. These patterns are again *opposite* to those suggested by self-control theory, as the theory postulates that those who *lack* self-control want stronger interventions as a commitment device. These interpretations, nevertheless, may be incorrect due to endogeneity issues (e.g., omitted variable bias), or heterogeneity, typical of cross-country analyses.

study uses a laboratory experimental method with neutral framing. Therefore, the findings can be applied to any social dilemma regarding the need for formal enforcement.

One advantage of using a laboratory experimental method is its control. It is possible to study people's preferences between formal and informal institutions without suffering from econometric issues. Using a carefully constructed design, this study provides the first experimental evidence that the state of people's self-regulatory resources influences their reliance on the formal enforcement of norms in a social dilemma. The self-regulatory resources of the recruited human subjects are manipulated using the crossing-out-letters (Baumeister *et al.*, 1998) and Stroop (1935) tasks. When their self-regulatory resources are *not depleted*, most decide not to introduce a costly formal sanctioning institution in a public goods game ("PGG," hereafter); however, they successfully cooperate through decentralized peer-to-peer punishment. In contrast, when they *are forced to deplete* their self-regulatory resources, the majority vote to implement a costly formal institution and then construct a deterrent punishment for free riders. Deterrent punishments have a strong effect on sustaining cooperation. Successful cooperation under the selected institutions is not due to the act of voting per se, but the effect of sanctioning institutions. In sum, these results emphasize that people's demands for formal sanctioning institutions depend on the amount of their self-regulatory resources and that they can overcome dilemmas by flexibly implementing appropriate institutions.

One remaining question is what the main driver of people's institutional preferences is, their self-control preferences (i.e., desire to use a commitment device), or their belief about peers' misbehavior under self-regulatory depletion. Another concern may be external validity owing to the neutral framing design of the laboratory experimental approach. To address these two points together, as a case study, the present study additionally conducts a survey (opinion) to supplement the main results by collecting information on people's self-control behaviors, beliefs about others' behaviors, and opinions on restriction measures during the Covid-19 crisis. The results confirmed that those who exhibited weaker self-restraint behavior during the Covid-19 pandemic prefer to rely more on formal restrictions and sanctioning institutions to deal with the cooperation problem during the pandemic, which is consistent with our observations in the laboratory experiment. Furthermore, their elicited institutional preferences were unaffected by their beliefs about how others behave during the pandemic. Therefore, the laboratory results are externally valid, and, at least for the context of present social dilemmas, self-control types, *not beliefs*, are the key drivers behind their institutional formation.

The rest of the paper proceeds as follows: Section 2 summarizes the related literature, and Section 3 describes the experimental design. Section 4 presents the hypotheses based on a theoretical

analysis, and Section 5 reports the experimental results. Section 6 briefly explains the supplementary survey results. Section 7 concludes.

2. Related Literature

Two branches of literature are closely related to the present study: (a) social dilemmas and endogenous choices of institutions, and (b) self-control and self-regulatory resources. The branch in (a) emanates from economic experimental research, whereas that in (b) arises from theoretical work, economic experiments, and laboratory studies in neighboring fields (e.g., psychology).

First, a large volume of experimental research examines not only the human behavioral tendency to cooperate but also institutions' ability to sustain cooperation in social dilemmas. People's social dilemma behavior is often studied using a PGG —the game adopted in this study. In a PGG, human subjects are randomly assigned to a group of N (where $N > 2$), are given endowments, and then simultaneously decide how much to contribute to their group. The parameters are set such that members' privately optimal contribution levels are smaller than the socially optimal level. The PGG emulates many social dilemma situations, such as whether to litter, comply with laws and ordinances, or follow norms such as recycling and fulfilling civic duties. Prior research indicates that the tension between cooperation and free riding is intense: while some try to cooperate with their peers, they learn to behave uncooperatively as they gain experience (e.g., Ledyard, 1995; Chaudhuri, 2011).

Two types of institutions can facilitate cooperation. The first type is to utilize a *centralized* or *formal* institution that aligns members' private interests with group interests using deterrent incentives (e.g., Falkinger *et al.*, 2000; Tyran and Feld, 2006; Kosfeld *et al.*, 2009; Putterman *et al.*, 2011; Traulsen *et al.*, 2012; Zhang *et al.*, 2014; Kamei *et at.*, 2015; Nicklisch *et al.*, 2016; Fehr and Williams, 2018; Kamei and Tabero, 2021). For example, in Tyran and Feld (2006), while subjects contributed only 30% of the endowment in a standard linear PGG with free riding being the strictly dominant strategy, they on average contributed 93% of the endowment when a deterrent penalty scheme changed the equilibrium behavior to full contribution. Similarly, Falkinger (1996) and Falkinger *et al.* (2000) theoretically and experimentally show that a redistribution mechanism (which taxes free riders while subsidizing contributors) can lead to almost full efficiency. Furthermore, given the option of constructing a mechanism, most subjects can build a deterrent mechanism, thereby achieving a Pareto-efficient outcome (e.g., Putterman *et al.*, 2011; Kamei *et al.*, 2015).[5]

An alternative to a centralized solution is to rely on decentralized mechanism, such as peer-to-

[5] Kamei and Tabero (2021) show that decision-making formats may affect their voting behavior. They find that as a decision-making unit, "teams" vote more efficiently than "individuals" to deter free riding.

peer punishment (e.g., Fehr and Gächter, 2000, 2002; Masclet *et al.*, 2003; Gürerk *et al.*, 2006; Herrmann *et al.*, 2008; Gächter *et al.*, 2008; Casari and Luini, 2009; Ertan *et al.*, 2009). The theoretical prediction of free-riding does not change when decentralized punishment is available, based on the agents' self-interest and common knowledge of rationality.[6] However, experiments have demonstrated that members' informal punishment strongly improves efficiency, for as long as the costs to the punishers are not too high (e.g., Anderson and Putterman, 2006; Nikiforakis and Normann, 2008) and the interactions are sufficiently long (e.g., Fehr and Gächter, 2000, 2002; Gächter *et al.*, 2008). Various sets of authors have experimentally examined the factors that may explain informal punishment activities. Their explorations have successfully found non-material motives, for example, negative emotions (e.g., de Quervain *et al.*, 2004), inequity aversion and beliefs in peers' punishment (e.g., Fischbacher and Gächter, 2010), a conditional willingness to punish (e.g., Kamei, 2014), enjoying punishment activities (e.g., Casari and Luini, 2009), and culture and nationality (e.g., Herrmann *et al.*, 2008). Interdependent preferences, such as inequity aversion (e.g., Fehr and Schmidt, 1999 and 2010) and reciprocity (e.g., Rabin 1993, Charness and Rabin 2000), can theoretically rationalize human punishment behavior and its behavioral effects.

Over the last decade, scholars have actively examined people's scheme preferences and the conditions under which groups enact formal rather than informal governance schemes (e.g., Traulsen *et al.*, 2012; Andreoni and Gee, 2012; Zhang *et al.*, 2014; Kamei *et at.*, 2015; Nicklisch *et al.*, 2016; Fehr and Williams, 2018; Kamei and Tabero, 2021). The findings suggest that groups delegate sanctioning power to a central authority by voting when formal schemes do not entail a high fixed (e.g., administrative) cost (e.g., Kamei *et at.*, 2015), when members' anti-social peer-to-peer punishment is more harmful than erroneous enforcement by the formal authority (e.g., Nicklisch *et al.*, 2016), and when members cannot reach a normative consensus regarding contribution behaviors in their group (Fehr and Williams, 2018). The endogenous selection of institutions has additional positive effects in fostering cooperation norms by not only allowing sorting (e.g., Dal Bó *et al.*, 2010; Dal Bó *et al.*, 2019) but also by directly influencing members' preferences for cooperation or providing them with an opportunity to signal (e.g., Tyran and Feld, 2006; Dal Bó *et al.*, 2010; Sutter *et al.*, 2010; Kamei, 2016, 2019). Despite numerous studies in this area, all prior experiments on institutions have been conducted without considering subjects' self-control capacities. This study proposes that the amount of people's self-regulatory resources influences their scheme choices and efficiency in a novel design that manipulates their regulatory resources.

[6] This is based on the logic of backward induction. No one inflicts punishment on their peers as doing so is costly. Anticipating this, no one contributes more than their privately optimal contribution levels in social dilemmas (Fehr and Gächter, 2000, 2002).

The second closely related area involves theoretical and experimental studies of self-control and temptation. For at least the last forty years, many scholars in psychology and its neighboring fields have consistently demonstrated that human self-regulatory recourses are limited, and therefore people tend to succumb to temptation when the resources are used up—a phenomenon called "self-regulatory depletion" (see, e.g., Baumeister *et al.* [1994, 2007] and Muraven and Baumeister [2000] for a survey). One important feature here is that self-regulatory resources are used to control and manage *all* kinds of urges and temptations; if a person uses regulatory resources to suppress some temptations in one dimension, they may not be able to resist temptations in other dimensions because the resources will have diminished.

Since around 2000, economists have followed scholars in these other fields on self-control research because of its significant importance in economics. Gul and Pesendorfer (2001, 2004) and many other prominent theorists made breakthroughs by formally modeling self-control behaviors and people's tendency to commit. In particular, Gul and Pesendorfer (2001) axiomatize self-control preferences by introducing a new axiom, "set betweenness," in an expected utility framework. Their representation theorem states that an agent incurs a self-control cost when choosing an action if other tempting options exist in the choice set. Thus, the agent has a *commitment* preference, i.e., they prefer to narrow their choice set in advance by removing tempting options from the menu. While the self-control theory by Gul and Pesendorfer (2001) provides dynamically consistent preferences and therefore does not explain the psychologists' idea of *limited* self-regulatory resources and depletion, its variant, that is, the addiction model (Gul and Pesendorfer, 2007; Kamei, 2012), and multi-self models (e.g., Ozdenoren *et al.*, 2011; Fudenberg and Levine, 2006), can explain self-regulatory depletion.

Experimental testing of human self-control behaviors and commitment preferences has recently been conducted in economics (Bucciola *et al.*, 2011; Burger *et al.*, 2011; Houser *et al.*, 2018; Toussaert, 2018; Kocher *et al.*, 2017). Houser *et al.* (2018) and Toussaert (2018) serve as direct tests for Gul and Pesendorfer's (2001) theory. Both experiments revealed self-control and commitment preferences among some individuals. First, Toussaert (2018) classified subjects who worked on a tedious task while facing temptation according to whether they wanted to eliminate the temptation. A quarter to a third of the subjects were classified as the "self-control type" (those who believed in their successful self-control without such elimination) and did resist the temptation. Second, Houser *et al.* (2018) let subjects decide whether to perform a real-effort task or surf the Internet, with the option of committing to working by paying a fee to eliminate the Internet surfing option. Some subjects used the costly commitment option. Bucciola *et al.* (2011) addressed the possibility of self-regulatory depletion. Bucciola *et al.* let children (aged 6 to 13) fold as many sheets as possible while including

the so-called "Marshmallow task" in the experiment. Their results showed that exposure to consumption temptations (e.g., snacks) significantly undermined younger children's productivity. This study differs from these prior experiments because it investigates people's commitment behavior through institutional formation in a *strategic* environment where two motives coexist: self-control and inequity aversion. In the context of this study, formal enforcement is linked to people's commitment preferences, as it makes free riding materially unbeneficial.

3. Experimental Design

The experiment is built on the framework of a finitely repeated linear PGG. Subjects are randomly assigned to a group of five, and the grouping remains the same throughout the experiment (partner matching). While a group size of four is more commonly used, a size of five is adopted here to avoid tied votes in institutional formation with majority voting. Each subject in a group has an endowment of 20 points in every period and simultaneously decides how many points to contribute to the public account. The marginal per capita return (MPCR) is 0.4. Subject *i* receives the following payoff in Period *t* when contributing $c_{i,t}$:

$$\pi_{i,t}(c_{i,t}) = 20 - c_{i,t} + r\sum_{j=1}^{5} c_{j,t}, \text{ where } r = 0.4. \qquad (1)$$

Four treatments are implemented as a 2×2 between-subjects design (Table 1). The between-subjects feature is important because subjects' experience in one environment may spill over to their behaviors in another environment—a phenomenon called "behavioral spill-over" (e.g., Kamei, 2016; Bednar *et al.*, 2012).[7] The first treatment dimension of the 2×2 design is the amount of subjects' self-regulatory resources. The second treatment dimension is whether subjects have an opportunity to enact sanctioning schemes by voting. In the treatments with institutional choices, groups decide which scheme to implement, either a formal sanctioning scheme ("FS," hereafter) or an informal sanctioning scheme ("IS," hereafter). Three additional treatments were also conducted in addition to the four treatments (the detail will be explained in Section 5.4).

Table 1: *Treatments*

Treatment	Voting	Self-regulatory resources	Part 2 Condition for PGG
[Main treatment:]			
No-N	No	Not depleted (Normal)	Six phases without sanction scheme
No-D	No	Depleted (Small)	Six phases without sanction scheme

[7] Most research in this area used a between-subjects design (e.g., Traulsen *et al.*, 2012; Andreoni and Gee, 2012; Zhang *et al.*, 2014; Kamei *et at.*, 2015; Nicklisch *et al.*, 2016; Fehr and Williams, 2018; Kamei and Tabero, 2021; Tyran and Feld, 2006; Dal Bó *et al.*, 2010; Sutter *et al.*, 2010; Kamei, 2016).

Voting-N	Yes	Not depleted (Normal)	Six phases each with FS or IS scheme
Voting-D	Yes	Depleted (Small)	Six phases each with FS or IS scheme
[Additional treatment:]			
Exo-IS-N	No	Not depleted (Normal)	Six phases with imposed IS scheme
Exo-FS-D	No	Depleted (Small)	Six phases with imposed FS scheme
Exo-IS-D	No	Depleted (Small)	Six phases with imposed IS scheme

All treatments comprise Parts 1 and 2. Part 1, also called Phase 1, is the same for the four treatments. In Part 1, subjects play the PGG described above four times in sequence without institutional choices. Subjects' payoffs in each period are calculated using Equation (1). This part plays a role in familiarizing subjects with peers' incentives to free ride (e.g., Ledyard, 1995; Chaudhuri, 2011). Part 2 has six phases (each comprising four periods) and differs by treatment. The six phases are called Phases 2 to 7. Having multiple phases allows us to examine how experience affects institutional choices.

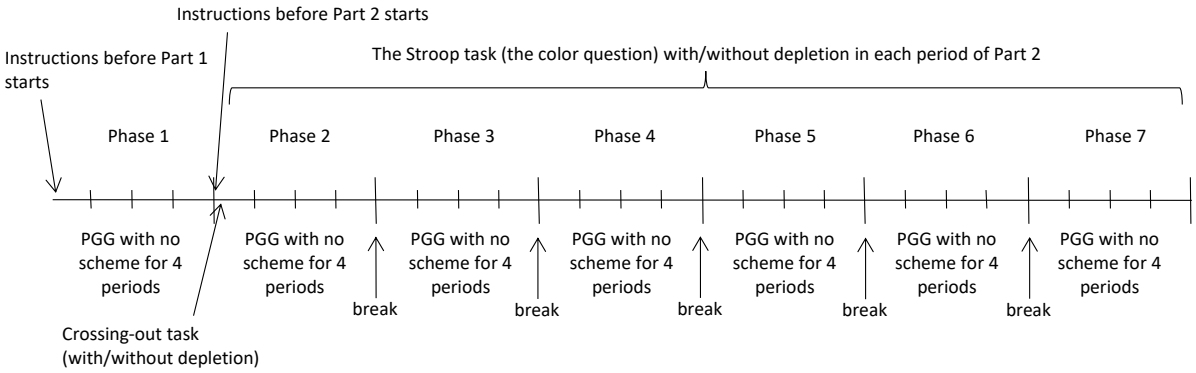
The two treatments without institutional choices are called the "No Voting, No Depletion" (No-N) and "No Voting, Depletion" (No-D) treatments. Part 2 of the No-N treatment begins with a task without depletion, followed by six phases, each with a four-period standard PGG. In contrast, subjects in the No-D treatment are *forced to deplete* their self-regulatory resources. Part 2 of the No-D treatment begins with a task parallel to the No-N treatment; however, the task contains an element that affects the self-regulatory resources (Subsection 2.1), whereafter the six phases of interactions commence. There is an additional depletion task in each period of Part 2 to maintain the manipulated state of self-regulatory resources throughout (Panel A of Figure 1).

The two treatments with institutional choices are called the "Voting, No Depletion" (Voting-N) and "Voting, Depletion" (Voting-D) treatments. The structures of the Voting-N and Voting-D treatments are the same as those of the No-N and No-D treatments, respectively, except for the opportunity to choose institutions (Panel B of Figure 1). The Voting-N (Voting-D) treatment has the same no-depletion (depletion) tasks as the No-N (No-D) treatment.

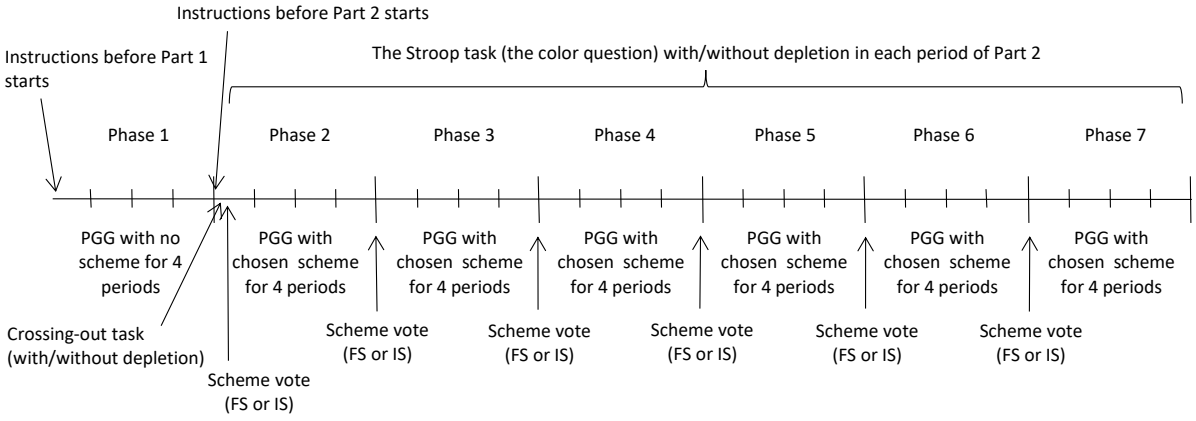
In Part 2 of the Voting-N and Voting-D treatments, subjects can use a sanctioning scheme in each period. The design of the institutional setting follows Kamei *et al.* (2015). At the onset of each phase, the groups can select an FS or IS by voting. Whichever scheme receives at least three votes (i.e., majority voting) will be in effect for the four periods in the given phase. The period structures vary according to the scheme. When a group selects the IS scheme, each period comprises two stages: allocation and informal (peer-to-peer) punishment. The first allocation stage is the same as in

Phase 1, in which each member decides how to contribute to the public account. When all the members have made their decisions, they will be informed of each member's contribution amount,

Figure 1: *Schematic Diagram*



(A) No-D and No-N treatments



(B) Voting-D and Voting-N treatments[#1]

Notes: [#1] When a group selects the FS scheme in a given phase, it decides a sanction rate by voting in each of the four periods in that phase. In other words, they have four voting opportunities in that phase.

and will then be allowed to assign punishment points to one another. These are costly punishment decisions; for each punishment point assigned to a member, one point is deducted from the punisher, while three points are deducted from the punished. There are two requirements here. First, the punishment points assigned to a member must be an integer. Second, punishment points must be less than or equal to 10 for any group member.

When a group selects the FS scheme, each period comprises two stages: a voting stage, and an allocation stage under the enacted FS scheme. Allocations to their private accounts are penalized in the FS scheme. The punishment strength is set to be equivalent to the IS scheme: the cost ratio is 1:3 (punisher: punished). At the beginning of each period, the members vote on the sanction rate to be

used. There are four possible rates: {0.0, 0.4, 0.8, 1.2}. The median of the five votes will be enacted in their group. The second stage is the allocation stage, as in each period of Part 1; however, it is subject to the FS scheme. There are the following two costs under the FS scheme. First, every subject must pay a fixed administrative cost of having the scheme (f = 5) in each period, irrespective of whether formal punishment is inflicted. The fixed cost can be thought of as a cost to eliminate the temptation to violate social norms. Using the two treatments, this study asks whether subjects prefer to commit to cooperation by enforcing the FS scheme and then selecting a deterrent sanction rate when self-regulatory ability is dominated by the size of free-riding temptations, as proposed by self-control theory.

The second is variable costs. For each point lost by a fined member, every group member incurs a cost of 1/11 points to impose that punishment. This is interpreted as the administrative cost of imposing the fine. Thus, the punished incurs a total loss of 12/11 (=1+1/11), while the four punishers incur a loss of 4/11(=4×1/11). The ultimate cost ratio is 3:1 (12/11: 4/11), the same punishment cost ratio as the IS scheme. Note, however, that these two schemes have different aspects. First, as already discussed, the FS scheme requires *fixed* cost payments. Second, punishment is only targeted at free riders with the strength collectively agreed in the FS scheme. In contrast, peer-to-peer punishment in the IS scheme depends on members' decisions; thus, it is possible that free riders may not be effectively punished and that high contributors may also be punished.

3.1. Depletion task

Two depletion tasks are used: one at the beginning of Part 2, and the other during the 24-period PGG of Part 2 (Figure 1). While the former task manipulates the amount of self-regulatory resources before PGG interactions begin in Part 2, the latter plays a role in maintaining the depleted state at low levels. The literature on self-regulatory resources (e.g., Baumeister *et al.*, 1994, 2007) states that people may recover from self-regulatory depletion and regain the ability to exercise self-control if a sufficiently long time passes after depletion or if they experience a positive mood. Thus, including the latter task can manipulate the amount of self-regulatory resources throughout Part 2.

This study uses the crossing-out-letters task ("crossing-out task," hereafter) to manipulate subjects' self-regulatory resources at the onset of Part 2. Hagger *et al.* (2010) performed a meta-analysis of the depletion tasks used in earlier experiments, suggesting that the crossing-out task is one of the most effective (Deng [2018] provided an updated meta-analysis). For example, Achtziger *et al.* (2016) and Gerhardt *et al.* (2017) used the crossing-out task following the suggestions by Hagger *et al*. (2010).

The subjects in the experiment perform the crossing-out task for eight minutes, although the

rule differs by treatment. The subjects in the no-depletion condition (the No-N and Voting-N treatments) cross out every letter *e* in a paragraph (from a well-known book) appearing on the computer screen and then submit the number of *e*'s. Once the subject submits an answer, a new paragraph appears on the computer screen. The paragraph is not short (Appendix A.3); thus, it is difficult to answer the question correctly. However, self-regulatory resources are not required because the task rule is simple. In contrast, the subjects in the depletion condition (the No-D and Voting-D treatments) cross out *e*'s, *except* if a vowel precedes it by two letters or if it is immediately followed by a vowel (the same rule was used in, e.g., Baumeister *et al.* [1998] and DeWall *et al.* [2011]). As in Baumeister *et al.* (1998), the paragraph flashes under the depletion condition, requiring extra attention (further depleting their mental resources). There are a maximum of six paragraphs. Subjects will be paid one point for each paragraph that they answer correctly. While making this task incentive-compatible is crucial for encouraging subjects to answer the questions seriously, leading to the successful manipulation of their self-regulatory resources, compensation is set at a minimum value to avoid the effects of receiving compensation (if any) on subsequent behaviors.

Further, a Stroop task (1935) is included in Part 2. In each allocation decision stage of Part 2, one of the four words ("red," "blue," "purple," and "black") randomly appears on the computer screen. In the depletion condition, the word has a color, either red, blue, purple, or black, while the color does *not* necessarily coincide with the word's meaning (e.g., the word "red" appears in blue). Moreover, the word flashes, thus affecting the subjects' attention. The coloring of the words is randomized. Subjects are asked to answer the color in which the word appears, along with the allocation decision in the PGG (see Appendix A for a screen image). For example, the answer is red if the word "blue" appears in red. By contrast, in the no-depletion condition, coloring always coincides with the meaning of the words (e.g., "red" appears in red), and the word does not flash.[8] A subject receives one point for each correct answer (the subject can earn up to 24 points, as there are 24 periods in Part 2).

4. Hypothesis

Standard theory prediction based on players' self-interest and common knowledge of rationality is straightforward because the experimental design uses a finitely repeated game. With the

[8] An alternative to the Stroop task could be an attention control task (e.g., Masicampo and Baumeister, 2008; DeWall *et al.*, 2011; Ainsworth *et al.*, 2014). In the attention control task, unrelated words, such as tree and forest, appear randomly for 10 seconds each on the subjects' computer screens. The subjects in the depletion condition are instructed not to see the words, while, in the no-depletion condition, the subjects are not given any instructions for the words. Implementing this task would be difficult because it is often not possible for researchers to judge whether subjects see the words. Therefore, the attention control task was not adopted in the present study.

logic of backward induction, no one would contribute any points to their public account in each period of the no-voting treatments since free riding is each player's strictly dominant strategy in the game ($\partial \pi_i / \partial c_i = -0.6 < 0$). Thus, complete free riding is the unique subgame perfect Nash Equilibrium in the No-N and No-D treatments. In equilibrium, each player receives a payoff of 20 points (= 20 + 0×5×0.4) per period. Having the IS scheme does not alter the free-riding equilibrium in the voting treatments since the standard theory predicts that no one will inflict punishment due to the cost (e.g., Fehr and Gächter, 2000, 2002).

However, the prediction under the FS scheme differs from that under the no-scheme condition or IS scheme (e.g., Falkinger *et al.*, 2000; Kamei and Putterman, 2015). When the FS scheme is in effect, it is materially beneficial for each player to vote for a sufficiently strong sanction rate, i.e., 0.8 or greater, so that contributing everything becomes the strictly dominant strategy (e.g., Putterman *et al.*, 2011). As stated above, the MPCR in the PGG is 0.4. While a median voting rule is used, the possibility of error (trembling hand perfection) encourages all members to vote for a deterrent rate because any one's vote can then be pivotal—see Selten (1975). By enacting a deterrent sanction rate, each player obtains a payoff of 35 points (= 0 + 20×5×0.4 – 5). This difference in equilibrium behavior implies that given the option to vote and the possibility that their votes are pivotal, all subjects will vote for the FS scheme with the aim of enforcing deterrent sanction rates after that.

4.1. The Self-Control Model

The present experiment manipulates subjects' self-regulatory resources by adopting the two depletion tasks. The effect of self-control can easily be incorporated into analysis using the well-known self-control preference model developed by Gul and Pesendorfer (2001, 2004). However, the self-control model itself does not change the standard theory predictions just discussed. To demonstrate this, assume the following utility functional form (Gul and Pesendorfer, 2001):

$$U_i(S) = max_{x_i \in A}[\pi_i(x_i) - f \cdot 1_{FS} - SR \cdot (20 - x_i) \cdot 1_{FS} - c_i(x_i)],$$

$$\text{where } c_i(x_i) = \max_{y \in A} v_i(y) - v_i(x_i) \equiv \rho_{i,s} \cdot \left(\max_{y \in A}[\pi_i(y) + SR \cdot y \cdot 1_{FS}] - \pi_i(x_i) - SR \cdot x_i \cdot 1_{FS}\right). \quad (2)$$

Here, i indexes individual players, $S \in \{FS, IS\}$, A is the choice set in the PGG, i.e., [0, 20], $\pi_i(x_i)$ is given by Equation (1), f is the fixed cost (= 5), SR is the sanction rate enacted in the group, and $1_{FS} = 1(0)$ when the FS (IS) scheme is chosen, i.e., $S = FS$ (IS). $v_i(x_i)$ is Player i's temptation utility and is defined as the material benefit of behaving selfishly. $c_i(x_i)$, called the self-control cost of Player i, is the size of the temptation Player i resists by exercising self-control. It is defined as the difference

between the maximum temptation (i.e., $\max_{y\in A} v_i(y)$) and the current temptation *i* enjoys (i.e., $v_i(x_i)$). $\rho_{i,s}$ indicates the depletion of *i*'s self-regulatory resources. This specific form of the self-control cost with $\rho_{i,s}$ was used in Kamei (2012). The subscript *s* in $\rho_{i,s}$ indicates whether the self-regulatory resources are depleted (*s* = *D*) or not depleted (s = *N*). As self-regulatory depletion renders a player susceptible to temptation, such depletion is modeled by allowing $\rho_{i,s}$ to enlarge such that:

$$\rho_{i,D} > \rho_{i,N}. \tag{3}$$

In this self-control framework, Player *i* behaves the same under the FS scheme as the standard theoretical prediction described above because *i* and their group members will vote to enact deterrent sanction rates (*SR* = 0.8 or 1.2) for material reasons. Individual interests are aligned with group interests in a deterrent FS scheme, by which they contribute full endowment amounts. In other words, the self-control cost is zero. Formally, *i*'s payoff is calculated as 40 – *f* when the FS scheme is in place:

$$U_i(FS) = \max_{x_i\in A}\left[\pi_i(x_i) - \rho_{i,s}(\pi_i(20) + 20SR - \pi_i(x_i) - x_i \cdot SR) - f - SR \cdot (20 - x_i)\right]$$
$$= \pi_i(20) - f = 40 - f.$$

By contrast, individual private interests conflict with group interests in the IS scheme. Player *i* incurs a self-control cost accordingly: $c_i(x_i) = \rho_{i,s} \cdot (\pi_i(0) - \pi_i(x_i))$. This cost strengthens their motivations to contribute nothing to their group (i.e., $x_i = 0$). In equilibrium, each player obtains 20 points as payoff in the IS scheme, as follows:

$$U_i(IS) = \max_{x_i\in A}\left[\pi_i(x_i) - \rho_{i,s}(\pi_i(0) - \pi_i(x_i))\right]$$
$$= \max_{x_i\in A}\left[\left(1 + \rho_{i,s}\right)\pi_i(x_i)\right] - \rho_{i,s}\pi_i(0) = \left(1 + \rho_{i,s}\right)\pi_i(0) - \rho_{i,s}\pi_i(0) = \pi_i(0) = 20.$$

In summary, the self-control model predicts that all groups in the two voting treatments choose the FS scheme, whereafter they enact deterrent sanction schemes by voting and then contribute the full endowment amounts in the allocation stage, as is the case for the standard theory prediction.

4.2. Incorporating Inequity-Averse Preferences in the Self-Control Model

However, over the last few decades, experiments have shown that *human* subjects behave differently from predictions based on self-interest and rationality. In particular, they can sustain cooperation with peers at a high level when peer-to-peer punishment is available (e.g., Fehr and Gächter, 2000, 2002). The positive effects of the IS scheme can be explained by other-regarding preferences (see, e.g., Fehr and Schmidt [2006] and Sobel [2005] for a survey). For example, Fehr and Schmidt (1999, 2010) mathematically proves that inequity-averse individuals can contribute

more with punishment than without punishment.[9] Prior experimental research on institutional choices further indicates that, given an option to vote, people choose the IS scheme rather than the FS and then sustain cooperation well under certain conditions, whose pattern is consistent with inequity aversion (e.g., Zhang *et al.*, 2014; Kamei *et al.*, 2015; Fehr and Williams, 2018; Kamei and Tabero, 2021).

As will be detailed in Section 5, in Part 1 of the experiment (i.e., without sanction schemes), subjects began contributing around 40% of the endowment, gradually decreasing contributions over time. This dynamic is consistent with the prevalence of social preferences (e.g., Fischbacher and Gächter, 2010). However, they can sustain high-level contributions once the IS scheme was implemented. Informal punishment was inflicted much more frequently on low rather than high contributors. As low contributors obtain larger payoffs than high contributors, inequity-averse members are more motivated to punish low contributors. In sum, the inequity-averse preference model can explain the behaviors of our subjects well.

As summarized in Proposition 1, groups' scheme choices are theoretically affected by members' states of self-regulatory resources (depleted or not) once their inequity-averse preferences are incorporated into the standard self-control model. The intuition behind this prediction is as follows. First, everyone contributes the full endowment without incurring any self-control cost when they collectively select the FS scheme, as then they enact a deterrent sanction rate. However, under the FS scheme, they must pay a fixed administrative cost per period. Second, members can likewise achieve high efficiency when they enact the IS scheme if they have sufficiently strong inequity-averse preferences. However, the members incur self-control costs because their private interests conflict with their group's interests. The self-control costs are higher when their self-regulatory resources are depleted than otherwise due to condition (3). Thus, to save on self-control costs, people are more likely to enact a FS scheme, albeit at a fixed cost, if their self-regulatory resources are depleted rather than not depleted. The following mathematically derives this prediction.

Assume the following utility functional form instead of Equation (2):

$$U_i(S) = max_{x_i \in A} \Big[\pi_i(x_i) - f \cdot 1_{FS} - SR \cdot (20 - x_i) \cdot 1_{FS} - c_i(x_i) - \mu_i \sum_{j \neq i} \big(\pi_i(x_i) + SR \cdot x_i \cdot 1_{FS} - \pi_j(x_j) - SR \cdot x_j \cdot 1_{FS}\big)^2\Big],$$

$$\text{where } c_i(x_i) = \rho_{i,s}\big(\max_{y \in A}[\pi_i(y) + SR \cdot y \cdot 1_{FS}] - \pi_i(x_i) - SR \cdot x_i \cdot 1_{FS}\big). \quad (4)$$

[9] The prevalence of inequity-averse preferences among people can also be seen in altruistic punishment by bystanders (e.g., Fehr and Fischbacher, 2004; Kamei, 2020).

Here, μ_i indicates Player i's utility weight on income inequality. Incorporating inequality concerns using a quadratic functional form is sometimes used for simplicity (e.g., Kamei 2018; Chen and Kamei 2018). Obviously, the inclusion of members' inequity aversion does not alter the prediction under the FS scheme because individual selfish motives and group interests are aligned thanks to deterrent sanctions; and $U_i(FS) = 40 - f$.

However, this alters the prediction under the IS scheme. For an illustrative purpose, the reminder considers the case of a symmetric contribution situation.[10]

The optimal behavior under the IS scheme is analyzed by finding the optimal control x_i, that is, the contribution level that maximizes the inside of the squared bracket in Equation (4). The first-order condition here is written as follows:

$$\frac{\partial U_i(IS)}{\partial x_i} = -1 + r - 2\mu_i(-1)\sum_{j\neq i}\left(\pi_i(x_i) - \pi_j(x_j)\right) + \rho_{i,s}(-1 + r)$$
$$= (-1 + r)\left(1 + \rho_{i,s}\right) + 2\mu_i \sum_{j\neq i}\left(-x_i + x_j\right). \qquad (5)$$

Suppose that all *j* but *i* choose c^* as their contribution amounts: $x_j = c^*$. Then, if the self-control cost is sufficiently small, i.e., $\rho_{i,s}$ is sufficiently small such that $\frac{2\mu_i(N-1)}{(1-r)} - 1 > \rho_{i,s}$, *i* also chooses $x_i = c^*$ as their optimal response; thus, $x_k = c^*$ for all *k* holds as an equilibrium outcome.[11] In equilibrium, $U_i(IS) = 20 + (1 - 0.6\rho_{i,s})c^*$.[12]

Whether the Pareto-efficient equilibrium ($c^* = 20$) maximizes their utility level under the IS scheme depends on the size of $\rho_{i,s}$, as what value of c^*maximizes U_i depends on $\rho_{i,s}$ as follows:

$$c^* = 20 \text{ if } \rho_{i,s} < 5/3; \text{ then } U_i(IS) = 40 - 12\rho_{i,s};$$
$$c^* = 0 \text{ if } \rho_{i,s} > 5/3; \text{ then } U_i(IS) = 20.$$

Comparing the above utility level in the IS scheme against that in the FS scheme, it is concluded that players prefer the FS (IS) scheme if they have sufficiently small (large) amounts of self-regulatory resources, i.e., $\rho_{i,s} > (<) 5/12$. Note that when $\rho_{i,s} > 5/3$, $U_i(IS) = 20 < U_i(FS) = 35$. Proposition 1 summarizes these analyses. Since the available amount of subjects' self-regulatory resources in the Voting-D treatment is small due to depletion, the FS scheme is predicted to be more prevalent in the Voting-D treatment than in the Voting-N treatment.

---

[10] In this situation, theoretically no one would inflict costly punishment as doing so just create income inequality among the members.

[11] If $x_i = c^* + 1$, the right-hand side of Equation (5) = $(-1 + r)\left(1 + \rho_{i,s}\right) - 2\mu_i(N - 1) < 0$. If $x_i = c^* - 1$, the right-hand side of Equation (5) = $(-1 + r)\left(1 + \rho_{i,s}\right) + 2\mu_i(N - 1) > 0$, provided that $\frac{2\mu_i(N-1)}{(1-r)} - 1 > \rho_{i,s}$.

[12] $U_i(IS) = \left(1 + \rho_{i,s}\right)\pi_i(c^*) - \rho_{i,s}\pi_i(0) = \left(1 + \rho_{i,s}\right)(20 - c^* + 0.4 \cdot 5 \cdot c^*) - \rho_{i,s}(20 + 0.4 \cdot 4 \cdot c^*) = 20 + (1 - 0.6\rho_{i,s})c^*$.

Proposition 1: *The smaller the amounts of self-regulatory resources people have, the more strongly they rely on law enforcement to avoid self-control costs. In the context of the present study, members vote for enacting the FS scheme more frequently in the Voting-D than in the Voting-N treatment.*

5. Experiment Results

The experimental sessions were conducted face-to-face at the Experimental Economics Laboratory, Research Institute of Socionetwork Strategies at Kansai University, in November and December 2020 and July and August 2021.[13] An additional experiment, as will be explained in Section 5.4, was also conducted there in December 2023 and January 2024. A total of 290 students there (45, 45, 40, and 45 subjects in the No-N, No-D, Voting-N, and Voting-D treatments, respectively, and 115 subjects in the additional sessions) were recruited through the ORSEE (developed by Greiner, 2015) and participated in the experiment. In addition, a separate opinion survey was conducted in July 2022 to enhance the external validity of the findings (see Section 6 for details).

Appendix Table B.1 reports the performance on the crossing-out and Stroop tasks. It shows that the average scores in both tasks were significantly better in the non-depletion treatments than in the depletion treatments. While this is the expected pattern, the scores in the two tasks are economically very similar for the four treatments. This is also expected. Recall that correctly answering the Stroop's (1935) color questions is not difficult even in the depletion condition, although additional effort and attention are required. Accurately counting the number of *e*'s in the crossing-out task is difficult even in the non-depletion condition since each paragraph is lengthy. This small difference in performance helps remove the possibility of wealth effects gained from the task as a confounding factor in the data analysis.

A summary of the main experiment results is given in Table 2:

(a) How subjects resolve the public goods dilemma differs markedly according to their amount of self-regulatory resources. More than 80% of groups enacted the IS scheme when their self-regulatory resources were not depleted. In contrast, when the resources were depleted, approximately 66.7% of groups enacted the FS scheme.

(b) Each sanction scheme significantly helped the subjects overcome the dilemma. When the FS scheme was enacted, 66% to 74% of groups implemented deterrent sanction rates and achieved almost full efficiency (although they could not achieve strong cooperation when they failed to implement deterrent rates). When the IS scheme was enacted, punishments were targeted much more

[13] This is a standard economic laboratory with three tall partitions in each desk: one for the front and two for the sides.

at free riders, thereby achieving higher contributions than in the no-voting treatments.

Table 2: *Summary of the Main Result*

Self-regulatory resources	No-voting treatment	Voting treatment			
	Average contribution[#1]	% of groups where FS was enacted	% of deterrent sanction rates enacted in FS	Average contribution[#1]	
				Under deterrent FS	Under IS
Depleted	6.45	66.7%	73.5%	19.8	13.1
Not depleted	5.23	16.7%	65.6%	19.9	11.8

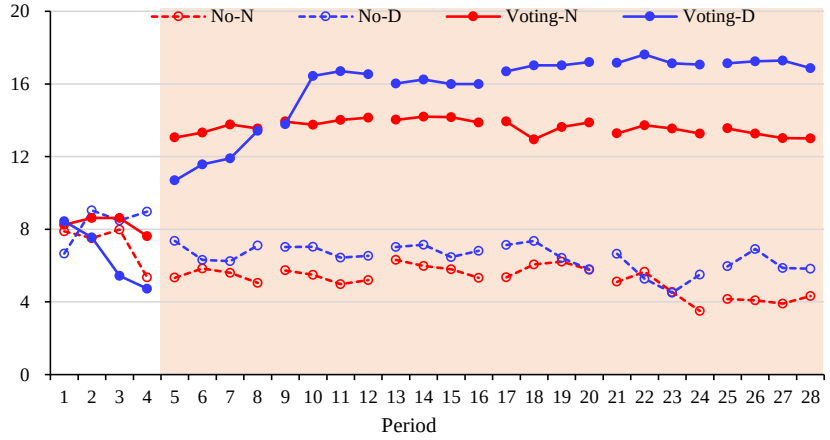
Note: [#1] The average contributions in Part 2 (periods 5 to 28).

The following subsections explain the details of the results: Section 5.1 briefly reports the contribution and payoff dynamics by treatment. Section 5.2 discusses subjects' scheme choices, after which Section 5.3 summarizes efficiency under enacted schemes. Section 5.4 reports additional experiments to explore whether voting per se affects behavior.

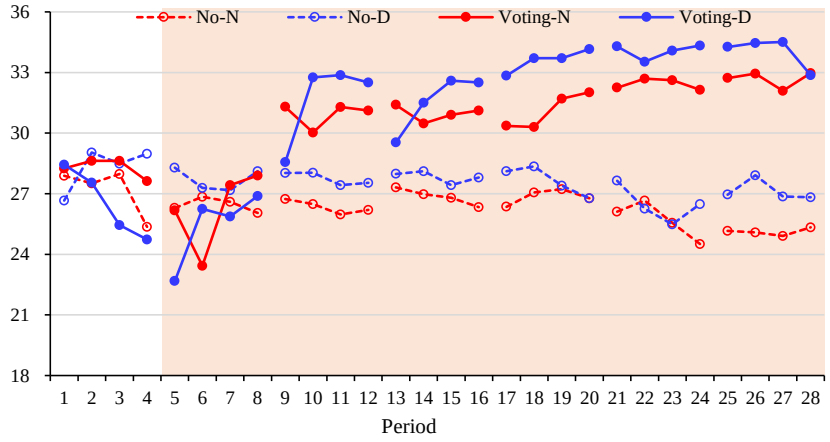
5.1. Contribution and Payoff

Figure 2 reports the contribution and payoff dynamics in each treatment. This indicates that the efficiencies were very similar in Part 1 (first four periods) for all four main treatments. For example, the average contribution was less than 50% of the endowment for each treatment. A Mann-Whitney test finds that the differences in the average contribution or payoff are not significant for any comparison (Panel I.i of Appendix Table B.2). It follows that the random assignment in the experiment was successful and that there was a large degree of free riding in each treatment without sanction schemes.

Figure 2: *Efficiency by Treatment*



(A) Average Contribution



(B) Average Payoff

Note: Period 1 to 4 are exogenous periods without sanction schemes.

The treatment effects of voting and self-regulatory resources on efficiency can be examined using the observations in Part 2. First, regardless of whether members' self-regulatory resources were depleted, free riding was serious when the sanction schemes were absent. Specifically, in the No-N and No-D treatments, the average contributions were consistently less than 40% of the endowment and the levels gradually declined over time. The difference in Part 2 average contribution was not significant between the two no-voting treatments (two-sided $p = 0.2475$, Mann-Whitney test). A regression analysis finds a qualitatively similar result (Appendix Table B.2). In the framework of the theoretical model (Section 4), this can be interpreted that without sanction schemes, the temptation to free ride is sufficiently strong in the public goods dilemma, irrespective of the amount of subjects' self-regulatory resources. Notice that if $\rho_{i,s}$ is sufficiently large, subjects cannot resist the temptation due to high self-control costs even if they are concerned about inequity. Floor effects are present as the subjects' contribution amounts are bounded below at zero.

Second, free riding was clearly deterred by the availability of sanction schemes (Panel I of Figure 2, Table 3). The effectiveness of punishment was not undermined by the state of members' self-regulatory resources. While in Part 1 the subjects in the Voting-N and Voting-D treatments experienced similar levels of free riding to those in the No-N and No-D treatments, the former achieved much higher contributions in Part 2, thereby receiving larger payoffs, than the latter. The average contribution or payoff difference is significant between the no-voting and voting treatments (Table 3). A regression analysis, whether a linear or tobit regression model is used, finds a qualitatively similar result (Part II of Appendix Table B.2). Thus, prior findings on the positive effects of punishment are robust to the state of people's self-regulatory resources.

Nonetheless, a close look at the data indicates that the impact of endogenous sanction schemes on payoffs is somewhat weaker than that on contributions owing to members' punishment loss. The impact is not significant when their resources are not depleted, that is, in the N-N versus Voting-N treatment (Panels I.iii and I.vi of Table B.2). The negative welfare effects of punishment are consistent with prior research: (a) punishment activities may be too intense under the IS scheme (e.g., Fehr and Gächter, 2000, 2002), and (b) a small number of groups may fail to construct a deterrent scheme and may therefore perform extremely poorly under the FS scheme (e.g., Group 13 of Putterman *et al.* [2011]). In contrast, these negative effects were somewhat milder in the Voting-D treatment. For example, a significantly larger percentage of groups in Voting-D still received a payoff of 30 points or greater compared with the No-D treatment. Here, the 30 points is the average payoff, assuming that the average contribution in a group is 50% of the endowment and no punishment is inflicted. As explained later, this difference in efficiency between the Voting-D and Voting-N treatments is driven mainly by the large difference in the scheme choice outcome: stronger popularity for the FS scheme in the Voting-D treatment than in the Voting-N treatment.

Result 1: *(i) Free riding was serious in both the No-D and No-N treatments where the sanction schemes were absent. (ii) Sanction schemes significantly improved cooperation regardless of whether the amounts of subjects' self-regulatory resources were small.*

Table 3: *Treatment Effects of Endogenous Schemes on Contributions and Payoffs*

	Treatment		Two-sided *p*-value for H_0: (i) = (ii)
	(i) No voting (No-D and No-N treatments)	(ii) Voting (Voting-D and Voting-N treatments)	
(a) Avg. contribution in Part 1[#1]	7.739 points (1.119 points)	7.362 points (0.965 points)	0.916
(b) Avg. contribution in Part 2	5.836 points (0.978 points)	14.809 points (1.575 points)	0.0001***
(c) Avg. payoff in Part 2	25.836 points (0.978 points)	30.259 points (1.711 points)	0.0349**
(d) % of groups whose Part 1 avg. contributions were ≥ 10 points[#1]	37.778%	38.529%	1.000
(e) % of groups whose Part 2 avg. contributions were ≥ 10 points	28.472%	76.912%	0.0005***
(f) % of groups whose Part 2 avg. payoffs were ≥ 30 points	27.454%	65.343%	0.0016***

Notes: *p*-values were calculated based on group-level Mann-Whitney tests for Rows (a) to (c) and Fisher exact tests for Rows (d) to (f). The numbers in parentheses in Rows (a) to (c) are standard errors clustered by group ID. No significant differences are found between the No-D and No-N treatments, as well as between the Voting-D and Voting-N treatments, in each of the six performance measures ((a) to (f)) – see Appendix Table B.2. [#1] The average

payoffs in Part 1 are monotonic transformations of the Part 1 average contributions based on Equation (1) for all treatments. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

5.2. Scheme Choice

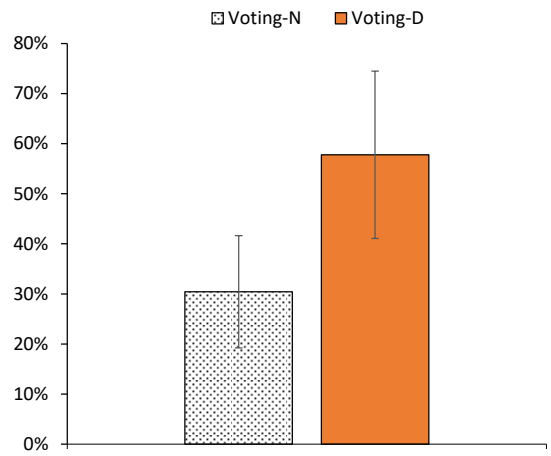
The popularity of the FS scheme and its vote outcomes differ markedly by the amount of self-regulatory resources. On the one hand, most depleted subjects preferred to use the FS scheme in the Voting-D treatment (Panel A of Figure 3). On the other hand, strikingly, only approximately 30% of non-depleted subjects voted for the FS scheme in the Voting-N treatment. These institutional preferences did not depend on the subjects' experience (see Appendix Figure B.3 for phase-by-phase voting). The difference in the vote share for the FS scheme is large: the vote share in the Voting-D treatment is approximately double that of the Voting-N treatment. This voting pattern is consistent with the prediction from the self-control and inequity-averse preference models summarized in Proposition 1.

Table 4 reports the results from the regression analysis of the treatment difference in the subjects' scheme votes. Model 1 of the table includes only the "Depleted" dummy (which equals 1 for the Voting-D treatment) to identify the treatment difference with a large dataset. It indicates that the depleted subjects' stronger preference for the FS scheme relative to that of the non-depleted subjects is significant. Model 2 includes available demographic variables as additional independent variables to control for possible differences in subjects' characteristics. It confirms that the impact of self-regulatory depletion remains significant by almost the same magnitude. Last, Models 3 and 4 add the vote number variable (which equals the phase number minus 1) and its interaction with "Depleted" to evaluate whether the effects of depletion vary as the subjects gain experience. The results show that both the vote number and the interaction are far from significant. This suggests that depleted (non-depleted) subjects' preferences for the FS (IS) scheme persist in the experiment.

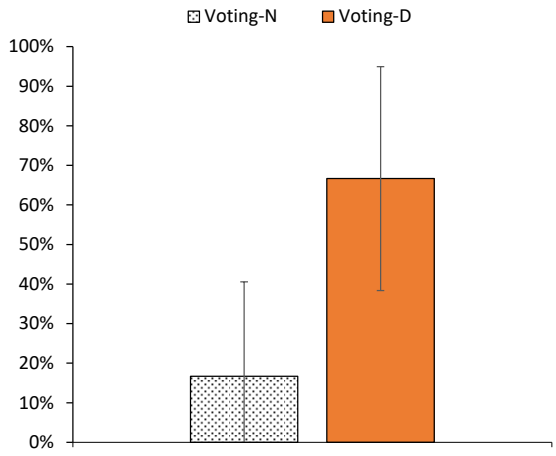
A majority rule was applied to determine a group's scheme. Panel B of Figure 3 reports the scheme choice outcomes by treatment. It indicates that the FS scheme was implemented much more frequently in the Voting-D treatment than in the Voting-N treatment. Regression analysis suggests that, parallel to the large differences in the popularity of sanction schemes (Panel A of Figure 3), the strong effect of self-regulatory resources on vote outcomes was significant (Panel B of Figure 3, Models 5 to 8 of Table 4).

Result 2: *Consistent with Proposition 1 of Section 4, subjects with smaller amounts of self-regulatory resources relied more on the FS scheme. Strikingly, the vote share of the FS scheme in the Voting-D treatment was approximately double that in the Voting-N treatment.*

Figure 3: *Popularity of the FS Scheme and Vote Outcomes*



(A) Popularity of the FS Scheme



(B) Scheme Choice Outcome (% in which the FS Scheme was Enacted)

Note: The error bars in each panel indicate the 95% confidence intervals (this applies to other figures).

Table 4: *Members' Amounts of Self-Regulatory Resources and Scheme Choices*

Dependent variable:	A dummy that equals 1(0) if Subject *i* voted for the FS (IS) scheme				The vote share of the FS scheme in Group $j \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$			
Estimation Method:	Subject random effects probit regressions with robust bootstrapped S.E. clustered by group ID.				Group random effects ordered probit regressions			
Independent variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(a) Depleted dummy {= 1 for the Voting-D treatment; 0 otherwise}	1.322*** (0.401)	1.241** (0.529)	1.787*** (0.664)	1.702** (0.712)	2.013*** (0.761)	2.036** (0.823)	2.937*** (0.925)	2.958*** (0.982)
(b) Vote number variable {= 1, 2, …, 6}	---	---	0.082 (0.074)	0.081 (0.082)	---	---	0.151 (0.096)	0.150 (0.096)
(c) Interaction (a) × (b)	---	---	-0.133 (0.120)	-0.132 (0.125)	---	---	-0.249* (0.131)	-0.248* (0.131)
(d) Constant	-0.968*** (0.289)	-1.065 (0.865)	-1.257*** (0.485)	-1.361 (0.996)	---[#2]	---	---	---
# of observations	510	510	510	510	102	102	102	102
Control variable[#1]	No	Yes	No	Yes	No	Yes	No	Yes
Wald χ^2	10.88	64.33	8.09	78.42	7.00	10.99	10.20	13.89
Prob > Wald χ^2	0.0010	0.0000	0.0443	0.0000	0.0082	0.0515	0.0170	0.0532

Notes: The numbers in parentheses are robust standard errors (S.E.). The units of observations are individuals in Models 1 to 4, and groups in Models 5 to 8. Group-level clustering was included in Models 1 to 4 as individuals' voting may be correlated within their group. Subject random effects linear regressions with robust standard errors (clustered by group ID) generate qualitatively similar results—see Appendix Table B.5.

[#1] The control variables include gender dummies, an economics major dummy, university year dummies, and political preferences in Models 2 and 4 [the percentage of female subjects, the percentage of economics majors, the

percentage of the first-year undergraduate students, and the average political preference in each group in Models 6 and 8]. [#2] The cut points were omitted to conserve space.
*, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

5.3. Performance Differences between the FS and IS Schemes

Both the FS and IS schemes are strong deterrents against free-riding.[14] Part I of Appendix Table B.4 reports the regression results examining how formal and informal punishment improved contributions relative to the no scheme condition of the no-voting treatments in Part 2. It shows that regardless of the scheme imposed, the average contribution was significantly higher in the Voting-D (Voting-N) treatment than in the No-D (No-N) treatment. The strong effects of the sanctioning schemes are not affected by the manipulation of self-regulatory resources (neither the interaction term between the Depleted and FS dummies nor that between the Depleted and IS dummies is significant). This again underlines the robustness of the punishment's role in social dilemmas.

Panel A of Figure 4 reports the contribution dynamics based on whether the sanction rate in the FS scheme was set at a deterrent level. It indicates that once a deterrent sanction rate was collectively enacted, the subjects contributed almost the full endowment irrespective of the state of their self-regulatory resources.[15] This contribution level was much higher than that under the IS scheme (see Models I and II of Table 5). However, this is not surprising as the full contribution is the unique Nash Equilibrium under the deterrent FS scheme. In contrast to the strong deterrence with high sanction rates, subjects failed to sustain cooperation when non-deterrent sanction rates were instead enacted (Table 5 and Panel A of Figure 4). This is also expected, as free riding is the unique Nash Equilibrium outcome with mild sanctions. On average, approximately 70% of the subjects voted for deterrent sanction rates in both the Voting-D and Voting-N treatments (Figure 4.C).

A somewhat interesting difference was observed between the states of self-regulatory resources: the average contribution under the IS scheme exhibited a significantly increasing trend in the Voting-D treatment, but not in the Voting-N treatment (Panel A of Figure 4, Part II of Appendix Table B.4). The same is observed for subjects' payoffs (Panel B of Figure 4). This difference in the efficiency trend can be explained by the significantly stronger pro-social punishment in the Voting-D treatment than in the Voting-N treatment (Panel D of Figure 4, Appendix Table B.8).[16] While pro-

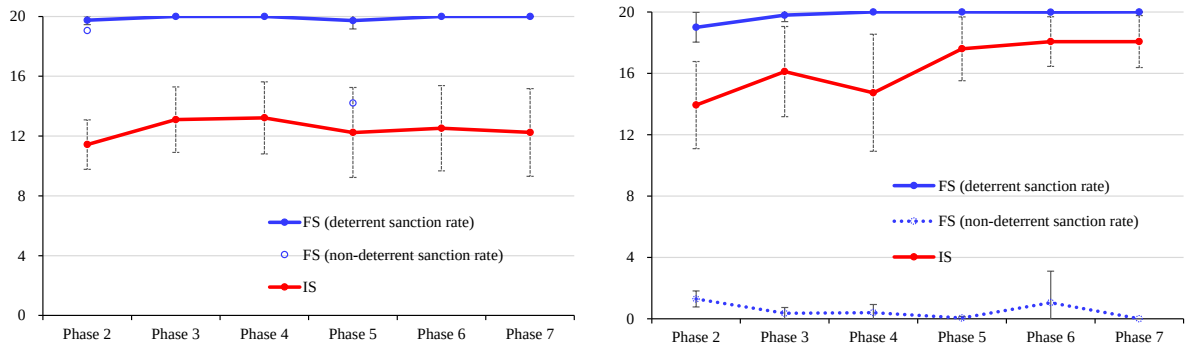
[14] The positive effect of the IS scheme, unlike Result 1.a, is consistent with the inequity-averse preference model. As discussed in Section 4.2, Fehr and Schmidt (1999) mathematically proved that contributions can be sustained at a high level more easily with than without peer-to-peer punishment.
[15] A regression analysis finds that the size of the sanction rate enacted in a group was a significantly positive predictor of the members' contribution amounts (Appendix Table B.6).
[16] Punishment received by Subject *i* in Period *t* is classified as pro-social (anti-social) when *i*'s contribution amount $c_{i,t}$ is *less than* (*not less than*) their group's average contribution amount $\bar{c_t}$ in Period *t*. This classification of pro-

social punishment was significantly stronger than anti-social punishment in both the Voting-D and Voting-N treatments (Appendix Table B.9), punishment activities among the *depleted* subjects were more intense, much more than double those among the non-depleted subjects (Panel D of Figure 4).[17] This feature fostered cooperation norms because those pro-socially punished in Period t increased their contribution amounts in Period $t + 1$, *ceteris paribus* (Appendix Table B.10).

Note that subjects have two conflicting sources of temptation in the punishment stage in the IS scheme: one is to free ride on peers' punishment acts, and the other is to inflict justice driven by negative emotions. The observed punishment patterns may mean that depleted individuals succumbed to *hot* temptation to respond to their negative emotions, rather than to *cool* temptation to free ride on others' punishment acts.[18] Of course, there are other possible interpretations; however, the bottom line here is that, despite the possibly better effects of the IS scheme in the Voting-D treatment, depleted subjects preferred to rely on the FS scheme.

Result 3: *(a) Voting on sanction rates and contribution decisions under the FS scheme was similar for the Voting-D and Voting-N treatments. In particular, deterrent FS schemes sustained contributions at almost the full efficiency level. (b) The IS scheme was effective in boosting contributions because pro-social punishment was stronger than anti-social punishment for both voting treatments.*

Figure 4: *Performances under Selected Schemes*



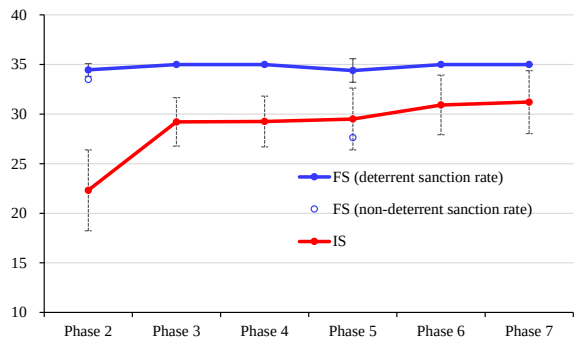
i. Voting-N treatment[#1] ii. Voting-D treatment

A. Average Contribution by Sanction Scheme

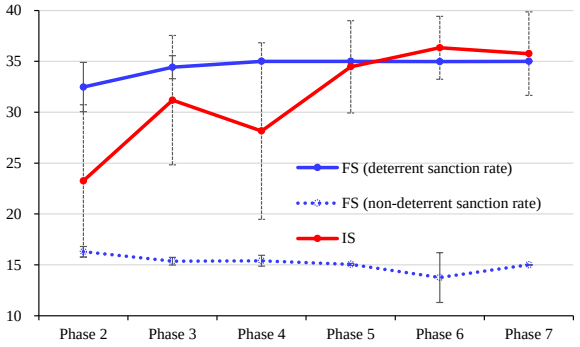
versus anti-social punishment is the same as what Fu and Putterman (2018) call the perverse versus the normal punishment.

[17] These punishment patterns hold for each of the six phases in Part 2 (Appendix Figure B.1).

[18] This interpretation resonates with the view that (a) punishment decisions may be driven by negative emotional states (e.g., de Quervain *et al.*, 2004), and (b) such motives may be hotter and stronger than their material motives to free ride on others' punishment.

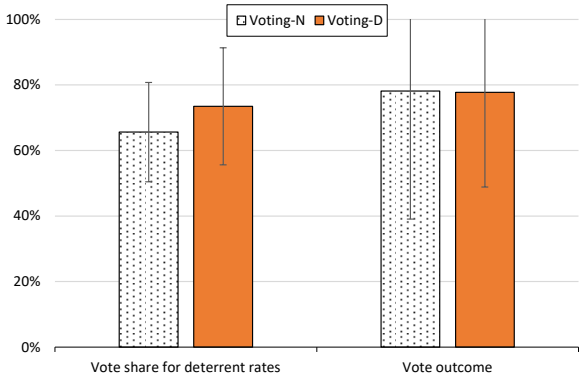


i. Voting-N treatment[#1]

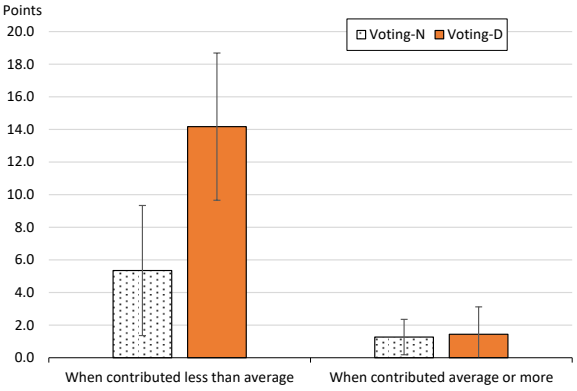


ii. Voting-D treatment

B. Average Payoff by Sanction Scheme



C. % of Voting on Deterrent Sanction Rates (≥ 0.8) under the FS scheme



D. Avg. per Period Loss due to Punishment Received under the IS scheme[#2]

Notes: [#1] The number of cases in which a group selected the FS scheme was much lower in the Voting-N treatment than in the Voting-D treatment. Non-deterrent rates were enacted only in Phases 2 and 5 in the Voting-N treatment. [#2] See Appendix Figure B.1 for the trend of average loss due to punishment, phase to phase. See Appendix Figure B.2 for the trend of average per subject punishment loss by sanction scheme.

Table 5: *Deterrence of the FS Scheme and Efficiency*

Dependent variable: / Independent variables:	Contribution of Subject *i* in Period *t*, where *t* > 4				Payoff of Subject *i* in Period *t*, where *t* > 4			
	I. Voting-D		II. Voting-N		III. Voting-D		IV. Voting-N	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
(a) Deterrent FS dummy	14.688***	14.781***	10.598***	11.066***	10.668***	11.277***	4.556*	3.626
	(1.338)	(1.227)	(2.505)	(2.444)	(1.803)	(1.894)	(2.636)	(2.654)
(b) Non-deterrent FS dummy	-3.504**	-3.477**	7.640***	8.115***	-6.569**	-6.586**	0.167	-0.806
	(1.742)	(1.652)	(2.732)	(2.564)	(2.882)	(2.838)	(2.852)	(2.693)
(c) IS dummy	6.965***	6.989***	8.084***	8.563***	-0.821	-0.856	4.683	3.715
	(2.427)	(2.366)	(2.759)	(2.593)	(4.746)	(4.592)	(2.861)	(2.714)
(d) Constant	6.447***	7.555***	5.225***	6.539***	26.447***	27.236***	25.225***	24.738***
	(1.142)	(1.798)	(1.565)	(2.948)	(1.142)	(2.095)	(1.565)	(3.122)
# of observations	2,160	2,160	2,040	2,040	2,160	2,160	2,040	2,040
Reference Group	No-D	No-D	No-N	No-N	No-D	No-D	No-N	No-N
Control variable[#1]	No	Yes	No	Yes	No	Yes	No	Yes
R-squared	0.5481	0.5598	0.2726	0.3062	0.2363	0.2626	0.0643	0.1195

Two-sided *p*-values for Wald test:								
H_0: (a) = (b)	0.0000***	0.0000***	0.0025***	0.0018***	0.0000***	0.0000***	0.0039***	0.0030***
H_0: (a) = (c)	0.0001***	0.0001***	0.0058***	0.0040***	0.0085***	0.0054***	0.9414	0.9583
H_0: (b) = (c)	0.0000***	0.0000***	0.1989	0.1836	0.0821*	0.0761*	0.0000***	0.0000***

Notes: Subject random effects linear regressions with robust standard errors (S.E.s) clustered by group ID. The numbers in parentheses are robust S.E.s. Observations from the No-D and Voting-D (No-N and Voting-N) treatments are used for Models I and III (II and IV). [#1] The control variables include gender dummies, an economics major dummy, university year dummies, and political preferences. *, **, and *** indicate significance at the 0.10, 0.05, and 0.01 levels, respectively.

5.4. Pure Effects of Voting

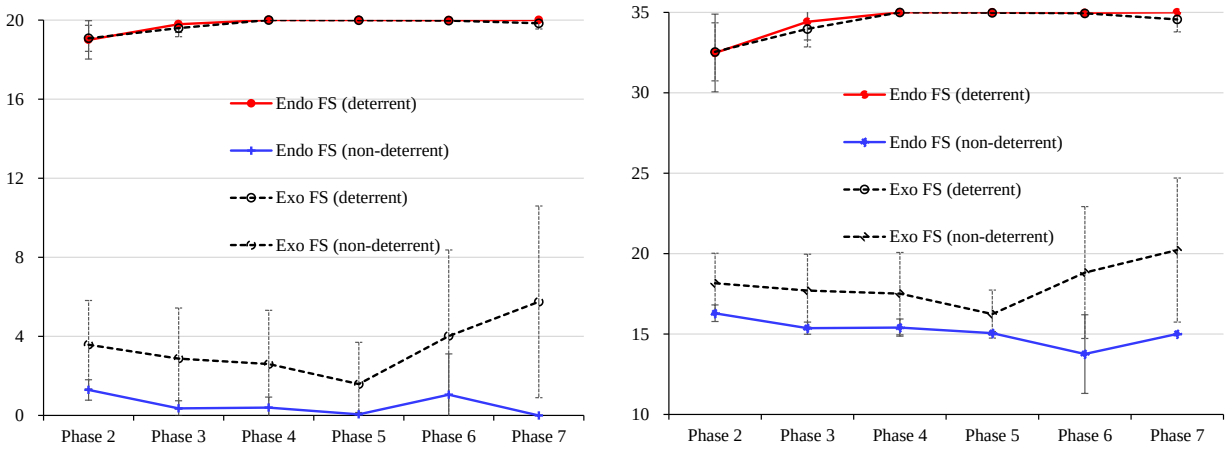
The FS scheme was enacted in the Voting-D treatment around 67% of the time (Figure 3.B). In this treatment, depleted subjects constructed deterrent FS schemes similar to those of non-depleted subjects in the Voting-N treatment (Figure 4.C), and achieved almost full efficiency under the enacted deterrent sanction rates (Figures 4.A and 4.B). The equally high performance in the Voting-D treatment (despite the subjects' self-regulatory depletion) might be due to the act of voting. For example, collectively deciding to pay a cost of five points per period to use the FS scheme may serve as a *costly signal* to commit cooperation (e.g., Tyran and Feld, 2006). In addition, voting per se may enhance pro-sociality among those involved (e.g., Dal Bó *et al.*, 2010; Kamei, 2016)—a phenomenon called the "democracy premium." The signaling effect and democracy premium may overcome the possible negative effects of self-regulatory depletion.

To explore whether depleted subjects can select appropriate sanction rates and then achieve high efficiency even without the act of voting, an additional treatment, named "Exo-FS-D" ("Exogenous FS, Depletion), was conducted. This treatment is identical to the Voting-D treatment except that the FS scheme was always implemented in each period of Part 2 without voting. In the FS scheme, five members of each group voted on which sanction rate to be used under the median voting rule. Fifty subjects (10 groups) participated in the additional treatment. In the Exo-FS-D treatment, subjects voted for deterrent sanction rates (0.8 or 1.2) 60.3% of the time. This percentage is somewhat lower than the Voting-D treatment (73.5%). Thus, the impact of voting is positive and around 13.2% (= 73.5% – 60.3%), but it is far from significant (two-sided $p = 0.584$).[19] Figure 5 summarizes the average contributions and payoffs in the Endogenous IS scheme under the Voting-D treatment and in the Exogenous IS scheme under the additional treatment. Strikingly, the groups

[19] This test is based on a subject random effects linear regression with robust standard errors clustered by group ID as the subjects made decisions repeatedly in their groups. Endogeneity has small effects (if any) in the present environment. A power analysis (power = 0.8, $\alpha = 0.05$, $\beta = 0.2$) shows that the required sample size to detect significance is 25.82 times as large as the data size in this study.

achieved almost full efficiency under deterrent sanction rates in both treatments. No significant differences in the efficiency measures were detected between the two treatments.[20] Similarly, strong free-riding prevailed in both treatments when the groups failed to enact a deterrent scheme. The absence of the effects of voting is not surprising, because deterrent formal mechanisms are known to affect behavior strongly even if enacted exogenously (e.g., Falkinger *et al.*, 2000; Putterman *et al.*, 2010). Therefore, the additional experiment confirms the high effectiveness of the FS scheme.

Figure 5: *Performance in the Endogenous vs. Exogenous FS scheme among Depleted Subjects*



A. Average Contribution

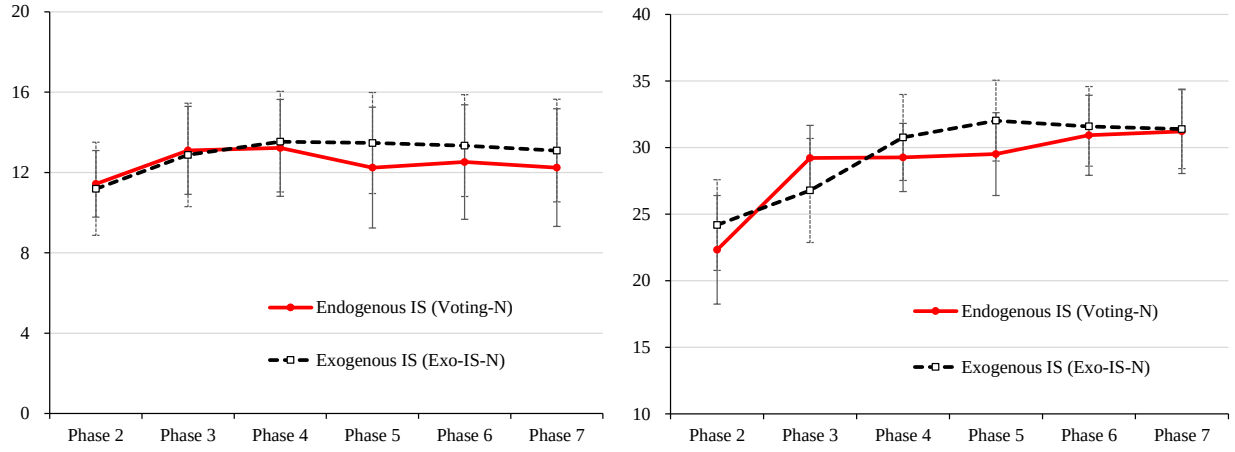
B. Average Payoff

In the Voting-N treatment, more than 80% of groups collectively implemented the IS scheme (Figure 3.B). The popularity of the IS scheme was significantly higher in the Voting-N than in the Voting-D treatment (Table 4). While this voting pattern was consistent with the theoretical prediction, the average contribution under the IS scheme was significantly lower than that in the deterrent FS scheme, regardless of whether self-regulatory resources were depleted (Figure 4 and Table 5). One may consider it to be driven by negative signaling effects: Voting for the IS scheme may signal that they want to avoid the cost of committing the full contribution. Notice that although constructing a deterrent sanction scheme almost assures full efficiency (Figure 4.A), a fixed administrative charge for using the formal scheme was not small (five points). Negative signaling effects may make maintaining contributions at a high level difficult. However, on the other hand, voting in itself may positively affect their pro-sociality thanks to the democracy premium as already discussed. Therefore, it is unclear whether voting per se affected group cooperation norms in the IS

[20] The difference in the average contribution (payoff) between the endogenous and exogenous deterrent FS schemes is not significant at two-sided $p = 0.170$ (0.164) according to a subject random effects linear regression with robust standard errors clustered by group ID and demographic controls.

scheme negatively or positively. To check how voting affected contribution and punishment behavior in the IS scheme, an additional treatment, named "Exo-IS-N" ("Exogenous IS, No Depletion), was conducted by newly recruiting forty subjects (8 groups). The Exo-IS-N treatment began with the same Part 1 as the other main treatments, after which it had six phases with the imposed IS scheme in Part 2. Figure 6 compares the performance of the Endogenous IS scheme under the Voting-N treatment with that of the Exogenous IS scheme under the Exo-IS-N treatment. The results reveal very similar behavioral patterns for the two treatments. The differences in each behavioral measure are never statistically significant.[21] This suggests that voting per se did not affect almost anything in the subjects' decisions. This underlines the robustness of the IS scheme in the institutional setup.[22]

Figure 6: *Performance in the Endogenous vs. Exogenous IS scheme among Non-Depleted Subjects*

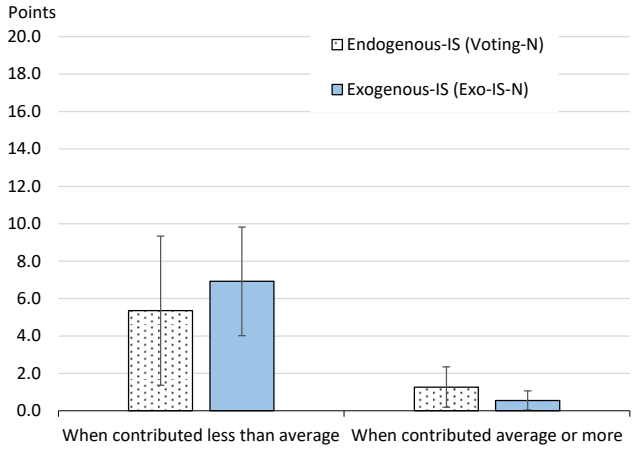


A. Average Contribution

B. Average Payoff

[21] The difference in the average contribution (payoff) between the endogenous and exogenous IS schemes is not significant at two-sided $p = 0.795$ (0.965) according to a subject random effects linear regression with robust standard errors clustered by group ID and demographic controls. The difference in the punishment received when contributed less than average (more than or equal to average) is not significant at two-sided $p = 0.584$ (0.452) according to a subject random effects linear regression with robust standard errors clustered by group ID and demographic controls. An alternative regression model (e.g., tobit) or a Mann-Whitney test generates qualitatively the same result for each comparison (the detail is omitted to conserve space).

[22] The efficiency of the IS scheme was somewhat higher in the Voting-D than in the Voting-N treatment (Panels A and B of Figure 4). As the IS scheme was enacted only 33% of the time (Panel B of Figure 3) in the Voting-D treatment, one may interpret this higher efficiency as in part driven by voting or selection bias. The number of groups that enacted the IS scheme at least once was five. To obtain deeper insights, the Exo-IS-D treatment was conducted by using new 25 subjects (five groups). As shown in Appendix Figure B.5, the average contribution and average payoff in the additional treatment are both similar to the ones in the IS scheme of the Voting-D treatment. This implies that the somewhat higher efficiency in the endogenous IS scheme among depleted subjects is driven by their self-regulatory depletion, not by the act of voting, as conjectured in Section 5.3.



C. Avg. Per Period Loss due to Punishment Received[#1]

Notes: [#1] The phase-by-phase trend of average loss due to punishment are also similar for the two treatments. See Appendix Figure B.4 for the detail.

Result 4: *Result 3 holds also for the FS and IS schemes exogenously implemented. Thus, results found in Section 5.3 were not driven by the act of voting.*

6. Supplementary Opinion Survey

While the laboratory experiment revealed the strong effects of people's self-regulatory resources on their policy preferences, to enhance the external validity of the findings, an opinion survey was conducted in July 2022 regarding people's self-control behaviors and policy preferences during the recent Covid-19 pandemic. A total of 298 students at Kansai University participated in the survey. The survey respondents were third- and fourth-year undergraduate students.[23]

A challenge in collecting information on self-control behaviors from respondents is the presence of a possible social desirability bias. Prior research has shown that people are reluctant to accept socially undesirable behaviors. For example, in an election, respondents are reluctant to accept their vote-buying experience (e.g., Gonzalez-Ocantos *et al.*, 2012). To avoid such bias, the respondents were provided with ten concrete examples, among which seven were on weak self-control behaviors (for example, "I saw my relatives (and/or your parents if you lived all by yourself), as normal. The frequency of seeing them was not much affected by the declaration of the state of emergency.") and three were on careful and high self-control behaviors (e.g., "I tried avoiding using

[23] Considering that the survey includes some questions on their behaviors as university students under the government's self-restraint requests, only third- or fourth-year undergraduates (as of July 2022) were recruited. Japan declared a state of emergency four times, and third- or fourth-year undergraduates experienced all the four self-restraint requests as university students.

public transportation (such as trains and buses) as much as possible."); the respondents were then asked to answer, in integers, the question of how many examples applied to their behaviors under the state of emergency (see Appendix C.1.1). There are two benefits of using this approach: First, it is possible to let respondents consider more concrete behaviors than when a question asks about their self-control abstractly (e.g., Did you comply with almost all the restriction measures imposed in the region?). This makes it possible to obtain a precise measure of self-control. Second, the ten examples include both socially desirable and undesirable behaviors, whose aspect makes it difficult for respondents to immediately notice what indicator the experimenter wants to see from the question. As seven (three) out of the ten examples were on weak (strong) self-restraint behavior, respondents' answers were expected to range from three to seven, such that a larger number would correspond to a weaker self-control type.[24]

The respondents were also asked a different question with ten examples, each of which described how the formal enforcement of restrictions could be strengthened (e.g., "The police should strengthen the patrol duties to monitor people's self-restraint behaviors during the periods when the government's request for self-restraint is in effect."); they were asked how many examples they agreed with—see Appendix C.1.3. The responses are used as the respondents' preferences for strong formal restrictions in the regression analysis.

In addition to these two key variables, the questionnaire asked about the respondents' perceptions of others' self-control (Appendix C.1.2). Not only people's preferences for strong formal restrictions and penalty, but also their self-control behaviors may be affected by their beliefs about others' behaviors, in which case an omitted variable bias may influence the result. An analysis of this elicited belief measure supplements the laboratory experiment, as beliefs were not elicited there. While the subjects in the experiment were more likely to support the FS scheme when their self-regulatory resources were depleted (Result 2), their voting may have been influenced by their beliefs about others' pro-social behavior without the scheme.

Appendix Section C.2 summarizes the results of the regression analysis. The results show that those who lack self-control to a larger degree are more likely to support strengthening the formal enforcement of restrictions. This significant correlation is not affected by whether people's perceptions of others' self-control or any other attribute variables are added. This result is also robust to the regression method used—a linear or tobit regression. The supplementary survey therefore helps rule out beliefs about others' behavior as a potential mechanism behind policy preferences and

[24] Indeed, most respondents selected numbers between 3 and 7. The percentages of those who selected the numbers 0, 1, 2, 8, 9, and 10 were 0.00%, 3.36%, 8.39%, 2.35%, 0.34%, and 1.01%, respectively.

confirms, in a realistic context, the key result of the significant relationship between people's self-control types and their commitment behaviors.

7. Conclusions

Social dilemmas are ubiquitous in both private and economic lives. During the last few decades, economic research has documented that dilemmas can be overcome by enforcing a formal institution that changes people's incentives. While a formal institution can effectively alter individuals' material interests such that these are aligned with their group's common interests, enacting it usually entails a cost, such as fixed administrative charges. Thus, the question remains open as to when implementing a formal institution is desirable, as groups can instead use decentralized peer-to-peer monitoring and punishment (e.g., Ostrom, 1990). This study is the first to show that the need for a centralized solution may depend on the state of the self-regulatory resources of a given group's members. In a novel laboratory experiment that manipulated their self-regulatory resources, most subjects preferred to govern themselves using monitoring and informal punishment when their resources were not depleted. However, when their resources were depleted, the majority of subjects preferred to rely on costly *formal* punishment. Their institutional choices were successful. The subjects achieved high group contributions under the respective sanction schemes. Moreover, a supplementary survey on the Covid-19 pandemic reveals that the findings from the laboratory are externally valid, and the main driver for their institutional preferences is their desire to commit cooperation, rather than their beliefs about peers' behavior.

This study is related to a large area of research on self-control and self-regulatory resources. The scheme choice preference found in the experiment is consistent with the self-control model formalized by Gul and Pesendorfer (2001, 2004) when combined with social preferences. This theory suggests that people with *small* amounts of self-regulatory resources incur a *high* self-control cost when they exercise self-control such that they do not succumb to the temptation to behave selfishly. The presence of disutility caused them to remove such temptations by voting in advance as a commitment device. By contrast, people can easily resist temptation when their resources are abundant. Thus, such strong self-control types sustain cooperation with informal punishment without relying on costly formal punishment. This experiment underlines the effects of self-regulatory resources on people's choices and the predictive power of the commitment theory in the context of endogenous institutional formation.

While the results obtained from the experiment are sufficiently clear, the present study is the first step in exploring the role of self-regulatory resources in an institutional setting. A meaningful further investigation is to sort out the underlying mechanism more precisely, for example by

identifying the impact of self-regulatory resources on preference parameters, such as ρ or μ. The present study adopted an approach similar to prior research, i.e., to measure the effects of depletion tasks on behavior, such as the impact of the temptation to surf the Internet (Houser *et al.*, 2018) or to read a story (Toussaert, 2018) on work behavior. However, the mechanism in social dilemma situations is more complex than the prior research in non-strategic settings, because two kinds of motives are involved: self-control and inequity aversion. Therefore, how people's relative importance of one motive to the other would look, and how the preferences differ by people, remains an open empirical question.

There are numerous research possibilities for verifying the robustness of the findings. For example, this study adopted experimental parameters frequently used in the experimental public goods literature (e.g., Ledyard, 1995; Chaudhuri, 2011), such as a group size of five, an MPCR of 0.4, and 24-period interactions in Part 2. However, there are other possible parameter values, such as different group sizes, in experiments. It is certainly a useful robustness test to study the same question by conducting experiments using different game parameters. Another example is that the accuracy of enforcement and noise may affect institutional formation. For simplicity, the present study assumes that not only do the subjects accurately observe their peers' contributions, but that punishments are also inflicted on the targets as intended. Such perfect observability and the absence of errors are typically assumed in the experimental literature for simplicity (e.g., Falkinger *et al.*, 2000; Tyran and Feld, 2006; Kosfeld *et al.*, 2009; Putterman *et al.*, 2011; Traulsen *et al.*, 2012; Zhang *et al.*, 2014; Kamei *et at.*, 2015; Fehr and Williams, 2018; Kamei and Tabero, 2021). However, Type I or II errors sometimes occur in real authorities or societies, which may affect institutional preferences (Nicklisch *et al.*, 2016).

Another important area for future research is to rule out possible alternative explanations behind people's institutional formation. For example, as an anonymous reviewer pointed out, the formal scheme differs from the informal scheme regarding the ease to use punishment. What subjects have to do under the formal sanctioning scheme is to select a sanction rate. If they successfully enacted a deterrent sanction rate and achieved high efficiency in some periods, they could repeat the same voting in later periods to receive high payoffs. In contrast, in the informal scheme, subjects must pay attention to others' contributions and judge how to punish their members in each period. Thus, arguably, the mental energy required for effective decision-making is larger in the informal scheme than in the formal scheme, and some depleted subjects might have just selected formal punishment to save energy, not to commit to cooperation. Further experiments would be useful to rule out this possibility, for instance, by making formal punishment as difficult as informal punishment.

Finally, care must be exercised when extrapolating any findings from one sample to another. Herrmann *et al.* (2008) demonstrated large differences in contribution and peer-to-peer punishment tendencies across cultures and societies. It is possible that there are cultural differences in the role of people's self-regulatory resources. For example, Herrmann *et al.* (2008) argued that the weaker the rule of law in a society is, the more prevalent anti-social punishment is. In a society where anti-social punishment is particularly widespread, people may always prefer a formal sanctioning scheme to avoid anti-social punishment, regardless of their self-regulatory resources. Conversely, they may prefer not to have a formal scheme because of their strong desire to punish others anti-socially. Which possibility is correct remains an open empirical question. Further experimental research would certainly be useful before the role of self-control can be generalized to the context of institutions and social dilemmas.

References

Achtziger, A., Alós-Ferrer, C., and Wagner, A. (2016), The Impact of Self-Control Depletion on Social Preferences in the Ultimatum Game. *Journal Economic Psychology* 53, 1-16.

Anderson, C., and Putterman, L. (2006), Do Non-strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism. *Games and Economic Behavior* 54(1), 1-24.

Andreoni, J., Gee, L. (2012), Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics* 96(11-12), 1036-1046.

Baumeister, R., Bratslavsky, E., Muraven, M., and Tice, D. (1998), Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology* 74(5), 1252-1265.

Baumeister, R., Heatherton, T., and Tice, D. (1994), *Losing Control: How And Why People Fail At Self-regulation*. San Diego, CA: Academic Press, Inc.

Baumeister, R., Vohs, K., and Tice, D. (2007), The strength model of self-control. *Current directions in psychological science* 16(6), 351-355.

Bednar, J., Chen, Y., Liu, T., and Page, S. (2012), Behavioral spillovers and cognitive load in multiple games: An experimental study. *Games and Economic Behavior* 74(1), 12-31.

Bucciola, A., Houser, D., Piovesan, M. (2011), Temptation and Productivity: A Field Experiment with Children. *Journal of Economic Behavior & Organization* 78, 126-136.

Casari, M., and Luini, L. (2009), Cooperation under alternative punishment institutions: An experiment. *Journal of Economic Behavior & Organization* 71(2), 273-282.

Charness, G., and Rabin, M. (2002), Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117(3), 817-869.

Chaudhuri, A. (2011), Sustaining Cooperation in Laboratory Public Goods Experiments: a Selective Survey of the Literature. *Experimental Economics* 14(1), 47-83.

Chen, J.I., Kamei, K. (2018), Disapproval Aversion or Inflated Inequity Acceptance? The Impact of Expressing Emotions in Ultimatum Bargaining. *Experimental Economics* 21, 836-857.

Dang, J. (2018), An Updated Meta-Analysis of the Ego Depletion Effect. *Psychological Research* 82, 645-651.

Dal Bó, P., Foster, A., and Putterman, L. (2010), Institutions and Behavior: Experimental Evidence on the Effects of Democracy. *American Economic Review* 100(5), 2205-2229.

Dal Bó, P., Foster, A., and Kamei, K. (2019), The Democracy Effect: a Weights-Based Identification Strategy. NEBR Working Paper 25724.

de Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., and Fehr, E. (2004), The Neural Basis of Altruistic Punishment. *Science*, 305(5688), 1254-1258.

Dekel, E., Lipman, B., and Rustichini, A. (2009), Temptation-Driven Preferences. *Review of Economic Studies* 76(3), 937-971.

DeWall, N., Mead, N., Baumeister, R., and Vohs, K. (2011), How Leaders Self-Regulate Their Task Performance: Evidence That Power Promotes Diligence, Depletion, and Disdain. *Journal of Personality and Social Psychology* 100(1), 47-65.

Ertan, A., Page, T., and Putterman, L. (2009), Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review* 53(5), 495-511.

Falkinger, J. (1996), Efficient private provision of public goods by rewarding deviations from average. *Journal of Public Economics* 62, 413-422.

Falkinger, J., Fehr, E., Gächter, S., and Winter-Ebmer, R. (2000), A Simple Mechanism for the Efficient Provision of Public Goods: Experimental Evidence. *American Economic Review* 90(1), 247-264.

Fehr, E., and Gächter, S. (2000), Cooperation and Punishment in Public Goods Experiments. *American Economic Review* 90(4), 980-994.

Fehr, E., and Gächter, S. (2002), Altruistic punishment in humans, *Nature*, 415(6868):137-140.

Fehr, E., and Schmidt, K. (1999), A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics* 114(3), 817-868.

Fehr, E., and Schmidt, K. (2010), On inequity aversion: A reply to Binmore and Shaked. *Journal of Economic Behavior & Organization* 73(1), 101-108.

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

Fehr, E, and Schmidt, K. (2006), The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories. In *Handbook of the Economics of Giving, Altruism and Reciprocity* by S.-G. Kolm and J. M. Ythier (eds.), 615-91. North Holland.

Fehr, E., and Williams, T. (2018), Social Norms, Endogenous Sorting and the Culture of Cooperation. IZA Discussion Papers 11457.

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics* 10, 171-178.

Fischbacher, U., and Gächter, S. (2010), Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review* 100(1), 541-56.

Fu, T., and Putterman, L. (2018), When is punishment harmful to cooperation? A note on antisocial and perverse punishment. *Journal of Economic Science Association* 4, 151-164.

Fudenberg, D., and Levine, D. (2006), A Dual-Self Model of Impulse Control. *American Economic Review* 96(5), 1449-1476.

Gächter, S., Renner, E., and Sefton, M. (2008), The long-run benefits of punishment. *Science* 322(5907), 1510-1510.

Gailliot, M. , and Baumeister, R. (2007). The Physiology of Willpower: Linking Blood Glucose to Self-control. *Personality and Social Psychology Review* 11, 303-327.

Gailliot, M., Baumeister, R., DeWall, C., Maner, J., Plant, E., Tice, D., Brewer, L., Schmeichel, B. (2007). Self-control Relies on Glucose as a Limited Energy Source: Willpower is More Than a Metaphor. *Journal of Personality and Social Psychology* 92, 325-336.

Gerhardt, H., Schildberg-Hörisch, H., and Willrodt, J. (2017), Does Self-Control Depletion Affect Risk Attitudes? *European Economic Review* 100, 463-487.

Gonzalez-Ocantos, E., de Jonge, C., Meléndez, C., Osorio, J., Nickerson, D. (2012), Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua, *American Journal of Political Science* 56, 202-217.

Greiner, B. (2015), Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association* 1(1), 114-125.

Gul, F., and Pesendorfer, W. (2001), Temptation and Self-Control. *Econometrica* 69(6), 1403-1435.

Gul, F., and Pesendorfer, W. (2004), Self-Control and the Theory of Consumption. *Econometrica* 72(1), 119-158.

Gul, F., and Pesendorfer, W. (2007), Harmful Addiction. *Review of Economic Studies* 74(1), 147-172.

Gürerk, Ö., Irlenbusch, B. and Rockenbach, B. (2006), The competitive advantage of sanctioning institutions. *Science* 312 (5770), 108-111.

Herrmann, B., Thöni, C., and Gächter, S. (2008), Antisocial punishment across societies. *Science* 319 (5868), 1362-1367.

Houser, D., Schunk, D., Winter, J., and Xiao, E. (2018), Temptation and Commitment in the Laboratory. *Games and Economic Behaviors* 107, 329-344.

Imperial College London, 2021. COVID-19 Global Behaviours and Attitudes. The Year in Review (April 2020 –April 2021), Institute of Global Health Innovation.

Kamei, K. (2011), Self-Regulatory Strength and Dynamic Optimal Purchase. *Economics Letters* 115(3), 452-454.

Kamei, K. (2014), Conditional Punishment. *Economics Letters* 124(2), 199-202.

Kamei, K. (2016), Democracy and Resilient Pro-Social Behavioral Change: An Experimental Study. *Social Choice and Welfare* 47, 359-378.

Kamei, K. (2018), Promoting Competition or Helping the Less Endowed? Distributional Preferences and Collective Institutional Choices under Intragroup Inequality. *Journal of Conflict Resolution* 62(3), 626-655.

Kamei, K. (2019), Cooperation and Endogenous Repetition in an Infinitely Repeated Social Dilemma. *International Journal of Game Theory* 48, 797-834.

Kamei, K. (2020), Group size effect and over-punishment in the case of third party enforcement of social norms. *Journal of Economic Behavior & Organization* 175, 395-412.

Kamei, K., Putterman, L., and Tyran, J.-R. (2015), State or Nature? Endogenous Formal versus Informal Sanctions in the Voluntary Provision of Public Goods. *Experimental Economics* 18, 38-65.

Kamei, K., and Tabero, K. (2021), The Individual-Team Discontinuity Effect on Institutional Choices. Working paper.

Kosfeld, M., Okada, A., and Riedl, A. (2009), Institution Formation in Public Goods Games. *American Economic Review* 99(4), 1335-55.

Kocher, M., Martinsson, P., Myrseth, K., Wollbrant, C. (2017), Strong, bold, and kind: self-control and cooperation in social dilemmas. *Experimental Economics* 20, 44-69.

Hagger, M., Wood, C., Stiff, C., and Chatzisarantis, N. (2010), Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis. *Psychological Bulletin* 136(4), 495-525.

Ledyard, J. (1995), Public Goods: A Survey of Experimental Research. In *The Handbook of Experimental Economics* by J. H. Kagel, A. E. Roth (eds.), 111-194, Princeton University Press.

Levitt, S. (1997), Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review* 87(3), 270-290.

Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press.

Ozdenoren, E., Salant, S., and Silverman, D. (2011), Willpower and the Optimal Control of Visceral Urges. *Journal of the European Economic Association* 10(2), 342-368.

Masclet, D., Noussair, C. Tucker, S., and Villeval, M.-C. (2003), Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review* 93(1), 366-380.

Mazar, N., Amir, O., Ariely, D. (2008), The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45(6),633-644.

Mead, N., Baumeister, R., Gino, F., Schweitzer, M., Ariely, D. (2009), Too Tired to Tell the Truth: Self-Control Resource Depletion and Dishonesty. *Journal of Experimental Social Psychology* 45(3), 594-597.

Muraven, M., and Baumeister, R. (2000), Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological bulletin* 126(2), 247-259.

Nicklisch, A., Grechenig, K., and Thöni, C. (2016), Information-sensitive Leviathans. *Journal of Public Economics* 144, 1-13.

Nikiforakis, N., and Normann, H.-T. (2008), A comparative statics analysis of punishment in public-good experiments. *Experimental Economics* 11, 358-369.

Putterman, L., Tyran, J.-R., and Kamei, K. (2011), Public goods and voting on formal sanction schemes. *Journal of Public Economics* 95(9-10), 1213-1222.

Sobel, J. (2005), Interdependent Preferences and Reciprocity. *Journal of Economic Literature* 43(2), 392-436.

Stroop, R. (1935), Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology* 18(6), 643–662.

Sutter, M., Haigner, S., and Kocher, M. (2010), Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Review of Economic Studies* 77, 1540-1566.

Toussaert, S. (2018), Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment. *Econometrica* 86(3), 859-889.

Traulsen, A., Röhl, T., and Milinski, M. (2012), An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proceedings of the Royal Society B* 279(1743), 3716-3721.

Tyran, J.-R., and Feld, L. (2006), Achieving Compliance when Legal Sanctions are Non-deterrent. *Scandinavian Journal of Economics* 108(1), 135-156.

Zhang, B., Li, C., De Silva, H., Bednarik, P., and Sigmund, K. (2014), The Evolution of Sanctioning Institutions: An Experimental Approach to the Social Contract. *Experimental Economics* 51(2), 285-303.

Vohs, K., Baumeister, R. (2016) *Handbook of Self-regulation: Research, Theory, and Applications*, The Guilford Press: New York.