Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

自己制御資源と罰則制度遂行：実験室内実験からの事実

亀井 憲樹

2022 年 9 月 26 日
2022-014
https://ies.keio.ac.jp/publications/19162/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
26 September, 2022

自己制御資源と罰則制度遂行：実験室内実験からの事実

亀井 憲樹

IES Keio 2022-014
2022 年 9 月 26 日
JEL Classification: C92, D72, H41
キーワード: 制度選択;社会的ジレンマ;公共財;セルフコントロール;罰則

【要旨】
本論文は、人々の自己制御資源の状態が社会的ジレンマ下におけるグループでの正式な罰則制度の遂行に影響を与えることを、実験室内実験の手法を用いて提案する。経済実験結果によると、自己制御資源が摩耗していない時には、過半数以上が（正式な罰則に頼らず）分権的なモニタリングとピア・ツー・ピアの罰則でジレンマを制御する決定をし、実際に高い協力を実現した。一方で、自己制御資源が摩耗し小さい場合には、過半数以上が、コストがかかるが規範逸脱者が自動的に罰則を受ける正式な罰則制度に投票・導入し、そのもとで、非協力者に対して抑止力のある強い罰則率を設定することで高い協力を実現した。外部妥当性を高めるため、新型コロナ危機に関するアンケート調査を補足として行ったが同様のパターンが観測された。これらの結果は不平等を嫌う選好とセルフ・コントロール選好から予測されるコミットメント行動によって説明できる。この行動経済理論によると、自己制御資源が小さい人は、コミットメント装置として、ただ乗りをする誘惑を投票を通じて事前に取り除くことで相互協力の実現を容易にし効用を高めようとする。本研究からの事実は、社会的ジレンマの文脈でのコミットメント行動の重要性を示していると解釈することができる。

亀井 憲樹
慶應義塾大学経済学部
〒108-8345
東京都港区三田2-15-45
kenju.kamei@keio.jp

謝辞：この論文は、独立行政法人経済産業研究所（RIETI）におけるプロジェクト「Self-regulatory resources and collective institutional choices in a social dilemma (an experimental study)」の成果の一部である。経済実験実施に際して科学技術融合財団から助成金（令和元年度調査研究助成）を受給した。また、経済実験室のコンピュータや実験ソフトウェアの準備・運営・管理、また被験者募集補助及び管理などロジスティックに関して、関西大学ソシオネットワーク戦略機構の小川一仁教授、元PD難波敏彦博士（現在京都先端科学大学講師）、嶋野温子様、大島敬恵様及び事務グループの方々から様々なサポートを頂いた。論文原案に対して、経済産業研究所ディスカッション・ペーパー検討会の方々から有益なコメントを頂いた。博士課程学生のMr. Artem Nesterov（ダラム大学）とMr. Xin Fang（早稲田大学）からzTreeに関するサポートを頂いた。科学技術融合財団、関西大学関係者、研究助手及びコメントを頂いた方に、ここに記して、感謝の意を表したい。

Self-Regulatory Resources and Institutional Formation: A First Experimental Test

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: This study conducts a novel laboratory experiment that shows 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 rigorously 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 amount of their resources is limited, the majority enact a costly formal sanctioning institution; backed by formal punishment, 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. 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. 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 the participants in the discussion paper seminar at RIETI.

1. Introduction

Human societies and organizations experience many 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 therefore 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 due to their self-control capacities (e.g., Gul and Pesendorfer, 2001, 2004; Baumeister *et al.*, 1994, 2007). Societies have ways to regulate opportunistic behavior through 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.

For the last 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 indeed induce real people to make socially optimal choices (e.g., Falkinger, 1996; Falkinger *et al.*, 2000; Putterman *et al.*, 2011). However, it at the same time, research shows that 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' choices between formal and informal sanctioning institutions, and have found that groups prefer to use a formal institution to a decentralized solution only under certain conditions, such as when the use of a formal institution does not entail a large 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). However, no studies have explored how people's self-control capacities, or more precisely, *self-regulatory resources*, are linked to their need for formal institutions in social dilemmas.

Self-regulatory resources are internal resources that people use to regulate their self-control, cope with stress and attention, and deal with conflicts between selfish and pro-social motivations. A large volume of experiments in neighboring fields of economics has consistently demonstrated, since around 1990, that (a) people's decision-making is strongly influenced by the state of their self-regulatory resources, and (b) the 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). 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 rigorously 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). While self-control and commitment theories may be applicable in the context of institutional formation in societies or organizations, surprisingly, this possibility has not been considered thus far.

How to implement a formal institution, such as formal punishment, is clearly a difficult but important issue. While such questions are ubiquitous and commonly raised in modern societies, identifying people's institutional preferences and the effects of policies is challenging. One example is restrictions related to the Covid-19 pandemic (which started 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 did resemble a self-regulatory depletion phenomenon. For example, Japan declared a state of emergency four times in the Tokyo area, and implemented strict restriction measures.[1] Based on data on people flow, however, the impacts of such restrictions diminished over time. For instance, on the first weekends following the declaration of the second, third, and fourth states of emergency, 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" (which means 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.[2] 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 such measures or government recommendations. They even tended to oppose restrictions when another wave later came. A survey, for example, showed that the percentage of those who were willing to self-isolate if advised decreased from 95% in April 2020 to 87% in April 2021 (Imperial College London, 2021). Parallel to people's attitudes, the country gradually shifted in the direction of 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 degrees of self-control are merely 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 the sizes of police forces in cities in the United States (US) (see Hoxby [2000] for another example). As policymaking in democratic countries such as Japan and the UK reflects people's views, the weaker restriction measures in

[1] 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.
[2] https://www5.cao.go.jp/keizai2/wellbeing/covid/pdf/result3_covid.pdf (in Japanese; accessed on Feb. 7, 2022)

a later Covid 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; 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 points of time less straightforward. People's concerns were not limited to their health and safety, as Covid-19 restrictions also both impacted labor markets and their 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 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, it is unclear who, those with strong or weak self-control, support such stronger formal enforcement. 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 country neighboring Japan or natural disaster.

A similar difficulty arises when this research question is examined based on an existing cross-country dataset. For example, the World Value Survey (WVS) Wave 7 (2017-2020) collected responses regarding what children were encouraged to learn at home, such as good manners and tolerance—see 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 WVS 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 correlation between the percentage of affirmative answers in Q7 to Q17 and the view on the government intervention was calculated as significantly negative (correlation = -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.[3] 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.

An advantage of using a laboratory experiment is its control. It is possible to study people's

[3] The percentage of the respondents' affirmative answers in Q7 to Q17 are significantly and negatively correlated with their average confidence levels on the police, courts, and government in the WVS (Q69, Q70, and Q71).

preferences between formal and informal institutions without suffering from econometric issues. With a carefully constructed design, this study provides the first experimental evidence that the state of people's self-regulatory resources does influence their reliance on the formal enforcement of norms in a social dilemma, as self-control and commitment theories, combined with inequity aversion, suggest. The recruited human subjects' self-regulatory resources are rigorously manipulated using two depletion/non-depletion tasks from the literature: the crossing-out-letters (Baumeister *et al.*, 1998) and Stroop (1997) 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 then successfully cooperate with one another through decentralized monitoring and peer-to-peer punishment. In contrast, when they *are forced to deplete* their self-regulatory resources, the vast majority vote to implement a costly formal institution, and then construct a deterrent punishment toward free riders. The deterrent punishment has a strong effect in sustaining cooperation. These results, therefore, emphasize that people's demands for formal sanctioning institutions in a social dilemma depend on the amount of their self-regulatory resources.

While the finding is quite convincing, one may be concerned about its external validity due to the neutral framing design of the laboratory experimental approach. Although the laboratory approach is standard and its usefulness is already well established, the present study additionally conducts a survey (opinion) to supplement the main experiment summarized above by collecting respondents' self-control behaviors and their opinions on the restriction measures during the Covid-19 crisis. The results obtained about their preferences in the field are consistent with our observations in the laboratory experiment: those who exhibit weaker self-restraint behavior during the Covid-19 pandemic prefer more to rely on formal restrictions and sanctioning institutions to deal with the cooperation problem during the pandemic.

The rest of the paper proceeds as follows: Section 2 briefly summarizes the related literature, while Section 3 describes the experimental design. Section 4 presents hypotheses based on theoretical analysis, while Section 5 reports the experimental results. Section 6 briefly explains the results of the supplementary survey. Section 7 concludes.

2. Related Literature

Two branches of the literature in the social sciences 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, while that in (b) arises from theoretical suggestions and experimental evidence in economics, as well as laboratory studies in neighboring fields such as psychology.

First, there is a large volume of experimental research that examines not only the human behavioral tendency to cooperate, but also institutions in sustaining cooperation in social dilemmas. People's social dilemma behavior is often studied using a PGG —the game adopted in this study, and among the most frequently used games in the literature. In a PGG, human subjects are randomly assigned to a group of N, where $N > 2$, are given endowments, and then simultaneously decide how many points to contribute to their group. Parameters are set such that members' privately optimal contribution levels are

smaller than the socially optimal level. The socially optimal contribution level is often set at the full endowment amount (i.e., linear public goods game); however, it is sometimes set at an interior level (i.e., non-linear public goods game). The PGG emulates many social dilemma situations, e.g., whether to litter, to comply with laws and ordinances, or to follow norms such as recycling and fulfil civic duties. Prior research indicates that tension between cooperation and free riding is intense. For instance, while some people attempt to cooperate with their peers, they learn to behave uncooperatively as they gain experience (e.g., Ledyard, 1995; Chaudhuri, 2011). Thus, some institutions are required to sustain cooperation, unless interactions are infinitely repeated.

There are two kinds of institutions that can facilitate cooperation. The first kind is to utilize a *centralized* or *formal* institution, which 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) showed, theoretically and experimentally, that a redistribution mechanism (which taxes free riders while subsidizing high contributors) can lead to almost full efficiency. Furthermore, given an option to construct a mechanism, most subjects can build a deterrent one, thereby achieving a Pareto-efficient outcome (e.g., Putterman *et al.*, 2011; Kamei *et al.*, 2015).[4]

An alternative to a centralized solution is to rely on decentralized, peer-to-peer monitoring and 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 standard theoretical prediction of free riding in PGGs does not change when decentralized punishment is available, based on agents' self-interest and common knowledge of rationality. 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 people's 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., Fehr and Fischbacher, 2004; 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., Hermann *et al.*, 2008). Interdependent preference models, 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. The high efficiency of decentralized solutions may mean

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

that groups do not need centralized solutions under certain conditions.

For the last decade, scholars have actively examined people's scheme preferences and the conditions under which groups enact formal, rather than informal, schemes for governance (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 do delegate sanctioning power to a central authority by voting when formal schemes do not entail a large 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 through voting (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 were conducted without considering the subjects' self-control capacities. The present study is, to the author's knowledge, the first to examine how the amount of people's self-regulatory resources influences their scheme choices and efficiency in a novel design that manipulates their regulatory resources in a laboratory.

The second closely related area involves theoretical and experimental studies on 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 the regulatory resources to suppress some temptations in one dimension, they may not be able to resist temptations in other dimensions since the resources will have diminished. The self-regulatory hypothesis is relevant for many economic transactions because people usually face conflicts in their economic decision-making: e.g., their individual decision-making, such as consumption choices and borrowing, and their social decision-making, such as whether to cooperate in a social dilemma, trust or betray others, etc.

Since around 2000, economists have followed scholars in these other social science fields on self-control research due to its significant importance in economics. Gul and Pesendorfer (2001, 2004) and many other prominent theorists first made breakthroughs by formally modeling human 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 in choosing an action if there are some other tempting options in the choice set (see also Dekel *et al.* [2009]). The agent, therefore, 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 psychologists' idea of *limited* self-

regulatory resources and depletion, its variant, i.e., 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 behavior and commitment preferences was conducted relatively recently 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 of Gul and Pesendorfer (2001)'s self-control theory. Indeed, both experiments revealed self-control and commitment preferences among some individuals. First, Toussaert (2018) found that self-control preferences were potentially dynamically consistent. In her experiment, subjects who worked on a tedious task while facing a temptation (i.e., to read a story during a task) were classified by 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 indeed resist the temptation during the task. A similar finding was obtained in an experiment by Kocher *et al.* (2017). Kocher *et al.* indicate that the stronger the self-control people have, the higher the level of cooperation they can achieve in a PGG. Second, Houser *et al.* (2018) let subjects decide whether to perform a real effort task ("counting" task) or surf the internet, with an option to commit to working by paying a fee to eliminate the internet surfing option. Some subjects did use the costly commitment option. Houser *et al.* (2018) also documented that self-control behavior might potentially be dynamically inconsistent when temptations were sufficiently strong,[5] as there were some subjects who delayed a commitment decision or succumbed to the temptation at a later stage. This is similar to the self-regulatory depletion phenomenon (Houser *et al.*'s experiment had a demanding, two-hour, task-solving task). The self-regulatory depletion possibility was carefully addressed by Bucciola *et al.* (2011). 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 result showed that exposure to consumption temptations (e.g., a snack) significantly undermined younger children's productivity. In the context of the present study, using formal enforcement is linked to people's commitment preferences in social dilemmas as it makes free riding materially unbeneficial. However, no study has investigated how people's self-regulatory resources influence their voting behavior and institutional formation outcomes. This study is the first to examine how the amount of people's self-regulatory resources affects their activation of formal enforcement, rather than their decentralized self-governance, in the context of endogenous institutional formation when there is tension between contributing and free riding. The amounts of subjects' self-regulatory resources are manipulated using well-established depletion tasks from the literature.

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 stays the same throughout the experiment (partner matching). Each subject in a group has an endowment of 20 points in every period, and then

[5] It is worth acknowledging that self-control behavior may be driven by a complex mechanism, as Burger *et al.* (2011) found that commitment devices might be counterproductive in the context of procrastination.

simultaneously decides how many points to allocate between their public and private accounts. The marginal per capita return (MPCR) is 0.4. In other words, Subject *i* receives the following payoff in Period *t* when they contribute $c_{i,t}$:

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

Four treatments are implemented as a 2×2 between-subjects design (Table 1); each subject plays the game under only one treatment condition. This 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).[6] 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).

Table 1: *Treatments*

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

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, as in Kamei *et al.* (2015). Subjects' payoffs in each period are calculated based on 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 (Periods 5 to 28) in the study. 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. By 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, whereafter the six phases of interactions commence (the depletion task will be explained in Subsection 2.1). There is an additional depletion task in each period of Part 2, to maintain the

[6] 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).

manipulated state of self-regulatory resources throughout. A schematic diagram of the two treatments is shown in Panel A of Figure 1.

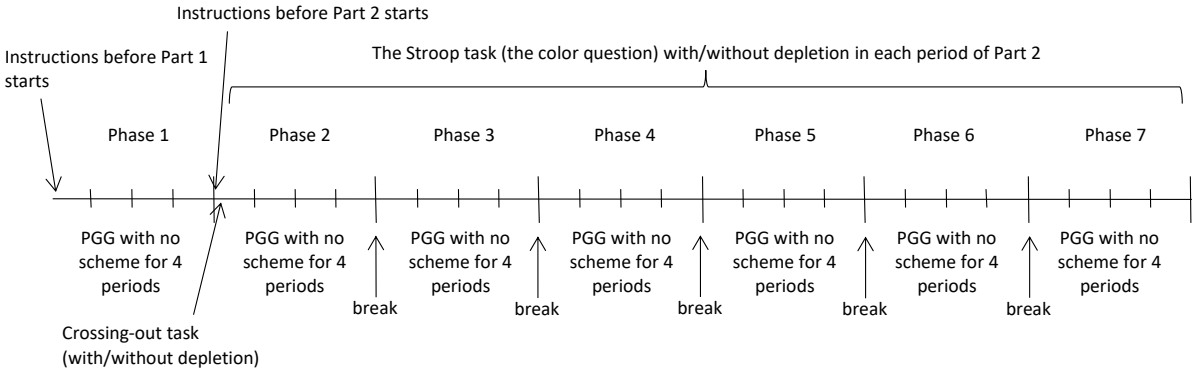
The two treatments with institutional choices are called the "Voting, No Depletion" (Voting-N) and "Voting, Depletion" (Voting-D) treatments. The structures of the Voting-N and Voting-D treatments are the same as those of the No-N and No-D treatments, respectively, except for the opportunity to choose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

In Part 2 of the Voting-N and Voting-D treatments, subjects can use a sanctioning scheme in each period of public goods interactions. The design for the institutional setting follows Kamei *et al.* (2015). At the onset of each phase, groups can select an FS or an 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. Voting is cost-free and mandatory. Period structures vary by scheme. When a group selects the IS scheme, each period comprises two stages: an allocation stage, and an informal (peer-to-peer) punishment stage. The first allocation stage is the same as the allocation stage in Phase 1: each member decides how to allocate 20 points between their private and public accounts. When all the members have made their allocation decisions, they will be informed of each member's contribution amount, and will then be provided with an opportunity 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 for the punishment decisions: First, the punishment points assigned to each member must be an integer. Second, the punishment points must be less than or equal to 10 for any one member of their group.

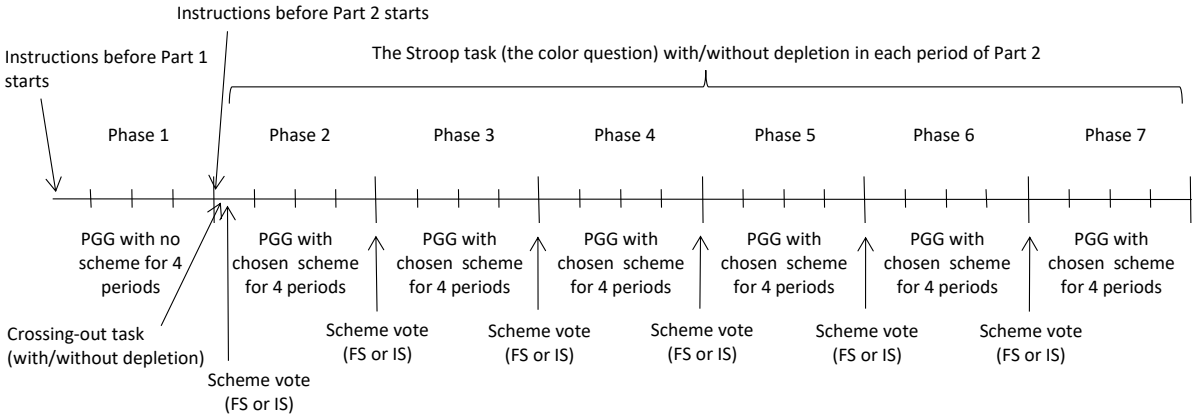
When a group selects the FS scheme, each period comprises two stages: a voting stage, and an allocation stage under the enacted FS scheme. Allocations to their private accounts are penalized in the FS scheme. The punishment strength is set such that it is equivalent to the IS scheme: the cost ratio is 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 decision stage as in each period of Part 1, but subject to the FS scheme. There are two costs under the FS scheme. First, every subject must pay an administrative fixed cost of having the scheme, $f = 5$, in each period, irrespective of whether formal punishment is inflicted (hence, the aggregate fixed cost per group is large, such that $5 \times 5 = 25$). This means that, while the comparative advantage of having an FS mechanism relative to an IS one is to enforce punishment precisely on free riders, it entails a large cost. The fixed cost can be thought of as a cost to eliminate the temptation to violate social norms in the public goods dilemma. Using the two treatments, this study asks whether subjects prefer to commit to cooperation by collectively selecting a deterrent sanction scheme when self-regulatory ability is dominated by the size of free riding temptations, as proposed by the self-control theory (e.g., Gul and Pesendorfer, 2001 and 2004). More specifically, this study asks, do subjects prefer using IS when they have sufficiently large self-regulatory resources? What happens to their choices between FS and IS when their resources have been depleted?

The second cost is variable costs. For each point lost by a member who is fined, every group member incurs a cost of 1/11 points to impose that punishment. This cost is interpreted as the administrative cost of imposing the fine. The punished thus incurs a loss of 12/11 (=1+1/11) in total, while the four punishers incur a loss of 4/11. The ultimate cost ratio is 3:1 (12/11: 4/11). In other words, the FS and IS schemes have the same punishment cost ratios. Note, however, that the two schemes have different aspects. First, as already discussed, the FS scheme requires *fixed* administrative cost payments, in contrast to the IS scheme. Second, punishment is only targeted at free riders with collectively agreed strength 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.

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.

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 (again, see Figure 1). While the former depletion task is used to manipulate the amount of self-regulatory resources before the public goods interactions begin in Part 2, the latter plays a role in

maintaining the depleted state at low levels. The literature states that people may recover from self-regulatory depletion and regain the ability to exercise self-control if a sufficiently long time passes after the depletion or if they experience a positive mood. For example, successful cooperation with deterrent punishment in Part 2 may help subjects recover their resources. Thus, including the latter task can manipulate the amount of self-regulatory resources throughout Part 2. Having the mental state depleted is also useful for real-world relevance in modeling people's smaller amounts of resources for a particular temptation in some societies (e.g., people may generally have smaller amounts of resources, and thus may tend to succumb to temptations such as littering [e.g., see the serious littering issue in the UK]; people may be tricked by moneylending businesses in a black market, which explains why strict regulations are required for some countries; addiction and drug use may also be related to self-regulatory resources, although these are more complicated due to the psychopathological symptoms that occur inside the brain).

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 depletion tasks in the literature, suggesting that the crossing-out task is one of the most effective (Deng [2018] provides 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).

In the experiment, the subjects perform the crossing-out task for eight minutes, although the rule differs by treatment. The subjects in the no-depletion condition (i.e., 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, one by one, and then submit the number of *e*'s. Once a subject submits an answer, a new paragraph appears on the computer screen. The paragraph is not short (see Appendix A.3), and thus it is not easy to answer the question correctly. However, self-regulatory resources are not required since the task rule is simple. By contrast, the subjects in the depletion condition (i.e., 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 in the depletion condition, thereby requiring extra attention by the subjects (and thus further depleting their mental resources). There are at most six paragraphs (one per screen) in this task. Subjects will be paid one point for each paragraph they answer correctly. While making this task incentive-compatible is crucial for encouraging subjects to seriously answer the questions (leading to successful manipulation of the self-regulatory resources), the compensation is set at a minimum value to avoid the effects of receiving compensation (if any) on subsequent behaviors.

Further, the so-called Stroop task (1992) is included during the public goods interactions to maintain the depleted state. In each allocation decision stage of Part 2, one of the four words ("red," "blue," "purple," and "black") randomly appears on the bottom of the computer screen. In the depletion condition, the word has a color, either red, blue, purple, or black, while the color does *not* necessary coincide with the meaning of the word (e.g., the word "red" appears in blue). Moreover, the word flashes, thus affecting subjects' attention. The coloring of the words is randomized. Subjects must answer in which color 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., the word "red" appears in red), and the word does not flash.[7] Thus, subjects can answer the color questions without using any self-regulatory resources. A subject receives one point for each correct answer in the Stroop task (the subject can earn up to 24 points as there are 24 periods in Part 2).

4. Hypothesis

The standard theory prediction based on players' self-interest and common knowledge of rationality is straightforward because the experiment design uses a finitely repeated game. With the 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' strictly dominant strategy in the game ($\partial \pi_i / \partial c_i = -0.6 < 0$). Thus, complete free riding is the unique sub-game perfect Nash Equilibrium in the No-N and No-D treatments. In equilibrium, each player receives a payoff of 20 (= 20 + 0 × 5 × 0.4) points 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 theoretical prediction under the FS scheme is different from that under the no-scheme condition or the IS scheme in the voting treatments (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 to their public account becomes the strictly dominant strategy (e.g., Putterman *et al.*, 2011). Recall that the MPCR in the PGG is 0.4. Note that while a median voting rule is used, the possibility of error (trembling-hand perfection) encourages all members to vote for a deterrent rate since any one's vote can then be pivotal—see Selten (1975). By enacting a deterrent sanction rate, each player obtains a payoff of 35 (= 0 + 20 × 5 × 0.4 – 5), rather than of 15 (= 20 + 0 × 5 ×0.4 – 5). This difference in the equilibrium behavior implies that, given an option to vote and the possibility that their votes are pivotal, all subjects would vote in favor of the FS scheme with the aim of enforcing deterrent sanction rates thereafter.

4.1. The Self-Control Model

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

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

[7] An alternative to the Stroop task could be an attention control task (e.g., Gilbert *et al.*, 1988; Masicampo and Baumeister, 2008; DeWall *et al.*, 2011; Ainsworth *et al.*, 2014). In the attention control task, (neutral) unrelated words, such as tree, forest, and water, appear randomly for 10 seconds each on the subjects' computer screens. The subjects in the depletion condition are instructed not to see the words and will be reminded during the experiment whether they see them, while, in the no-depletion condition, the subjects are not given any instructions for the words. Implementing this task would be more difficult than using the Stroop task because it is often difficult for experimenters to judge whether subjects see the words during the experiment. Therefore, the attention control task was not adopted in the present study.

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

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 administrative cost (= 5), *SR* is the sanction rate enacted in the group, and $1_{FS} = 1(0)$ when the FS (IS) scheme is chosen. $c_i(x_i)$ is Player *i*'s self-control cost and $\rho_{i,s}$ indicates the state of *i*'s self-regulatory resources. The specific form of the self-control cost with $\rho_{i,s}$ was used in Kamei (2012), and $\rho_{i,s}$ is called the "temptation index." The subscript, *s*, in $\rho_{i,s}$ indicates the state of self-regulatory resources, i.e., *D* (Depleted) or *N* (Non-depleted). As self-regulatory depletion renders a player susceptible to temptation, such depletion is modelled by allowing the temptation index to enlarge (see also Ozdenoren *et al.* [2011]). Therefore, in this theoretical framework, it can reasonably be assumed that:

$$\rho_{i,D} > \rho_{i,N}. \quad (3)$$

In this self-control framework, Player *i* behaves the same under the FS scheme as the standard theoretical prediction described above. Notice that *i* and their group members would vote to enact deterrent sanction rates (*SR* = 0.8 or 1.2) for material reasons in the FS scheme. This means that individual interests are aligned with group interests in that scheme, by which they contribute full endowment amounts. Thus, the self-control cost is zero in a deterrent FS scheme. More formally, Player *i*'s payoff is calculated as 40 – *f*:

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

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

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

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

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

For the last few decades, however, experiments have shown that *human* subjects behave differently from predictions based on self-interest and rationality (see Section 2). Especially, they can sustain cooperation with peers at a high level when peer-to-peer punishment is available (e.g., Fehr and Gächter, 2000, 2002), for as long as the punishment costs to the punishers are not too large (e.g., Anderson and Putterman, 2006; Nikiforakis and Normann, 2008). The positive effects of the IS scheme

can be explained by other-regarding preference models (see, e.g., Fehr and Schmidt [2006] and Sobel [2005] for a survey). For example, the inequity-averse preference model (Fehr and Schmidt, 1999, 2010) can successfully rationalize members' punishment behaviors and reactions to punishment received in social dilemma settings. 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). Prior research on institutional choices further indicates that, given an option to vote, people do 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).

This subsection shows that groups' scheme choices are theoretically affected by the members' states of self-regulatory resources, once their inequity-averse preferences are incorporated in the self-control model. Assume the following utility functional form, instead of Equation (2):

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

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

Here, μ_i is Player i's utility weight on income inequality.[8] Obviously, the inclusion of members' inequity aversion does not alter the prediction under the FS scheme, because individual and group interests are aligned with deterrent sanction rates; and $U_i(FS) = 40 - f$. Thus, the rest focuses on an analysis under the IS scheme while considering, for an illustrative purpose, a symmetric contribution situation. The assumption of symmetry significantly simplifies the analysis because no punishment is expected due to members' inequity concerns.[9]

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

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

$$= (-1 + r)\big(1 + \rho_{i,s}\big) + 2\mu_i \sum_{j \neq i}\big(-x_i + x_j\big). \quad (5)$$

Suppose that all *j* but *i* choose c^* as their contribution amounts: $x_j = c^*$. Then, if $\rho_{i,s}$ 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).[10] In equilibrium, $U_i(IS) = \big(1 + \rho_{i,s}\big)\pi_i(c^*) - \rho_{i,s}\pi_i(0) = \big(1 + \rho_{i,s}\big)(20 - c^* + 0.4 \cdot 5 \cdot c^*) - \rho_{i,s}(20 + 0.4 \cdot 4 \cdot c^*) = 20 + (1 - 0.6\rho_{i,s})c^*$. Thus,

[8] This quadratic functional form was used in Kamei (2018)'s theoretical analysis.

[9] The strategic situation is the one with multiple equilibria when other-regarding motives are added to the model, thus making the analysis quite complex if we also consider the cases with asymmetric equilibria.

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

whether the Pareto-dominant equilibrium ($c^* = 20$) also maximizes their utility level depends on the size of $\rho_{i,s}$, as the c^*that maximizes U_i depends on $\rho_{i,s}$ as follows:

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

$$\text{if } \rho_{i,s} > 5/3, c^* = 0 \text{ maximizes } U_i \text{ and } U_i(IS) = 20.$$

Combined with the optimal behaviors in the FS scheme, it can be concluded that players prefer the FS (IS) scheme if they have sufficiently small (large) amounts of self-regulatory resources, i.e., $40 - f > (<) 40 - 12\rho_{i,s}$, meaning that $\rho_{i,s} > (<) 5/12$. Note that when $\rho_{i,s} > 5/3$, $U_i(IS) = 20$, which is always less than $U_i(FS) = 35$. These analyses can be summarized in Proposition 1 as the main hypothesis of this study. Since the amount of subjects' self-regulatory resources in the Voting-D treatment is small, the FS scheme is predicted to be more prevalent in the Voting-D treatment than in the Voting-N one.

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

5. Experiment Results

The experiment 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.[11] A total of 175 students (45, 45, 40, and 45 subjects in the No-N, No-D, Voting-N, and Voting-D treatments, respectively), recruited through the ORSEE (developed by Greiner, 2015), participated in the experiment. No subjects participated in more than one session.

Appendix Table B.1 reports the performances in the crossing-out and Stroop tasks. It shows that the average scores in both the crossing-out and Stroop tasks are significantly better in the non-depletion than in the depletion treatments. While this is the expected pattern, however, the scores in the two tasks are economically very similar for the four treatments. This is also an expected result. Recall that answering the color questions correctly is not difficult even in the depletion condition, although additional effort and attention are required; moreover, 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 (Appendix A.3). This helps remove the possibility of wealth effects gained from the task as a confounding factor in examining the effects of the manipulation on subjects' institutional formation.

This section first describes the treatment differences in contribution and payoffs (Section 5.1), whereafter it examines subjects' scheme choice behaviors (Section 5.2). Lastly, their behavior under the enacted schemes is examined (Section 5.3).

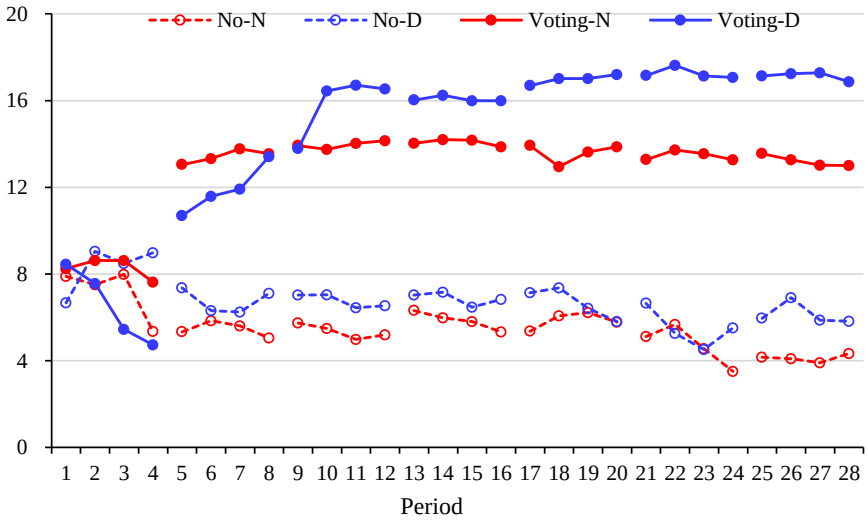
5.1. Contribution and Payoff

Figure 2 reports the contribution and payoff dynamics in each treatment. It shows that the

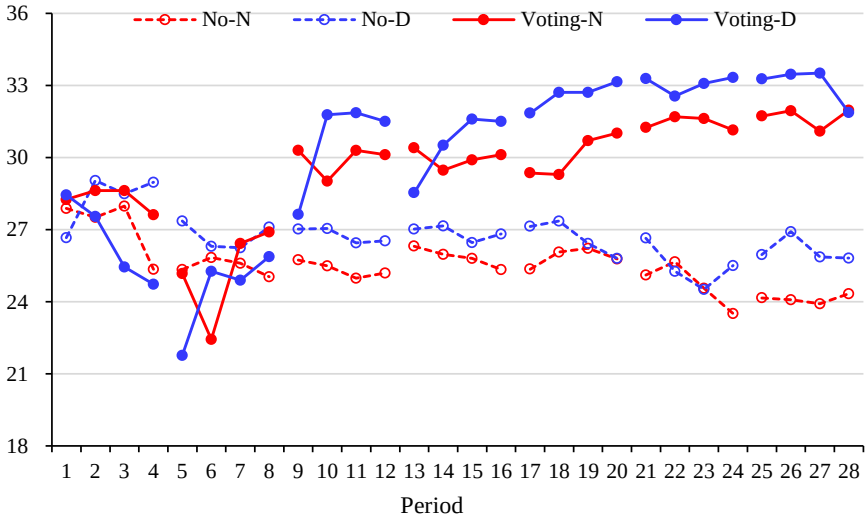
[11] This is a standard laboratory with three tall partitions in each desk: one for the front and two for the sides.

efficiencies are very similar in Part 1 for all four treatments. For example, the average contribution is less than 50% of the endowment in 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

The treatment effects of voting and self-regulatory resources on efficiency can be examined using the observations in Part 2. It shows first that, regardless of whether members' self-regulatory resources were depleted, free riding was serious when the sanction schemes were absent. Specifically, in both the No-N and No-D treatments, the average contributions were consistently less than 40% of the endowment, while the levels gradually declined over time. The difference in the average contribution in Part 2 is not

significant between the two no-voting treatments (two-sided $p = 0.2475$, Mann-Whitney test). A regression analysis finds a qualitatively similar result – see Part II of Appendix Table B.2.

Second, free riding was clearly deterred by the availability of sanction schemes (Panel I of Figure 2, Table 2). 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 levels of contributions in Part 2, thereby receiving larger payoffs, than the latter. The difference in the level of the average contribution or the average payoff is significant between the no-voting and voting treatments (Table 2 – see Part I of Appendix Table B.2 for more detailed results). A regression analysis, whether a linear or tobit regression model is used, finds a qualitatively similar result (Part II of Appendix Table B.2). This suggests that prior findings on the strong positive effects of punishment are robust to the amount of people's self-regulatory resources.

Nonetheless, a close look at the data indicates that the impact of voting on payoffs is somewhat weaker than that on contributions due to members' punishment loss. The impact is not significant for the 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, such negative effects seem to be milder for the Voting-D treatment. For example, a significantly larger percentage of groups in Voting-D still received 30 points or greater as a payoff, 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 punishments are inflicted. As will be explained later, this difference in efficiency between the Voting-D and Voting-N treatments is driven by (a) the large difference in the scheme choice outcome, and (b) the significantly stronger informal punishment activities seen in the Voting-D treatment relative to 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) Voting on sanction schemes significantly improved cooperation regardless of whether the amounts of subjects' self-regulatory resources were small. The positive effects of voting were somewhat stronger in the Voting-D treatment than in the Voting-N treatment, nevertheless.*

Table 2: *Treatment Effects of Voting on Contributions and Payoffs*

	(i) No voting[#1]	(ii) Voting[#2]	Two-sided p-value for H_0: (i) = (ii)
(a) Avg. contribution in Part 1[#3]	7.739 points	7.362 points	0.916
(b) Avg. contribution in Part 2	5.836 points	14.809 points	0.0001***
(c) Avg. payoff in Part 2	25.836 points	30.259 points	0.0349**
(d) % of groups whose Part 1 avg. contributions were ≥ 10 points[#3]	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). 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] "No voting" includes the No-D and No-N treatments. [#2] "Voting" includes the Voting-D and Voting-N treatments. [#3] 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

In contrast to the similar efficiencies in the two voting treatments, however, the popularity of the FS scheme and its vote outcomes differ markedly by the amount of self-regulatory resources. On the one hand, the majority of depleted subjects consistently 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 remained quite stable even after the subjects gained experience. The difference in the vote share for the FS scheme is large: the vote share in the Voting-D treatment is approximately double that in the Voting-N treatment. This voting pattern is indeed consistent with the prediction from the self-control and inequity-averse preference models summarized in Proposition 1.

Table 3 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 [0] for the Voting-D [Voting-N] treatment) to identify the treatment difference using all observations. It indicates that the depleted subjects' stronger preference for the FS scheme, relative to that of the non-depleted subjects, is strongly significant. Model 2 includes available demographic variables as additional independent variables to control for possible differences in subjects' individual 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.

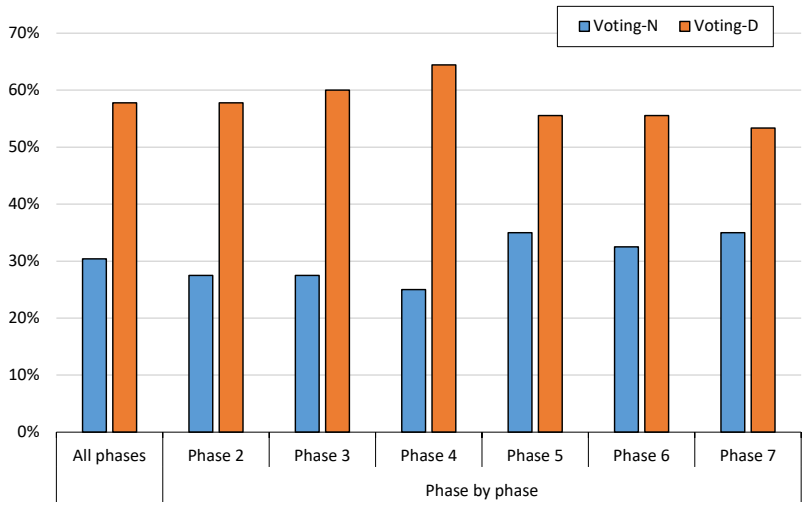
A majority rule was applied in the experiment to determine a group's scheme. Panel B of Figure 3 reports the scheme choice outcomes by treatment. It indicates that the FS scheme was implemented much more frequently in the Voting-D treatment than in the Voting-N treatment. Regression analysis suggests that parallel to the sustained differences in the popularity of sanction schemes (Panel A of Figure 3), the strong effect of self-regulatory resources on voting was not only significant, but it also persisted from phase to phase (Panel B of Figure 3, Models 5 to 8 of Table 3).

It is worth noting here that the difference in the scheme choice outcome (Panel B) is much larger than that in the scheme vote share (Panel A). This is due to the so-called "behavioral public choice theorem"—a key feature of the majority voting rule (e.g., Ertan *et al.*, 2009; Hauser *et al.*, 2014). Under majority voting, minorities tend to be outnumbered by the majority, thus allowing the latter's preference

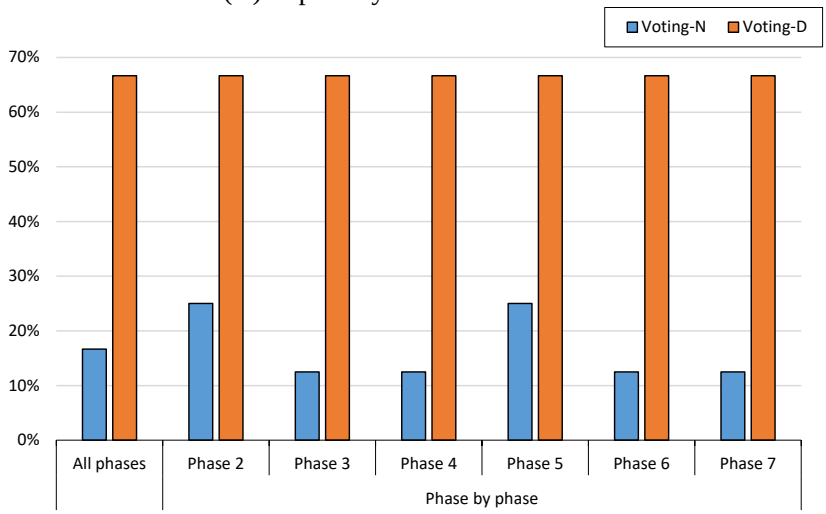
to be more easily enacted in the group.

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)

Table 3: *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 each individual's 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 a given 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, consistent with prior research. Part I of Appendix Table B.4 reports the regression results of examining how formal and informal punishment improved contributions in Part 2 relative to the no scheme condition of the no-voting treatments. 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 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 role of punishment in social dilemmas.

Panel A of Figure 4 reports the contribution dynamics based on whether the sanction rate in the FS scheme is 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.[12] The contribution level was much higher than that under the IS scheme (see Models I and II of Table 4). However, this is not a surprise as full contribution is the unique Nash Equilibrium under the deterrent FS scheme. In contrast to the strong deterrence with high sanction rates, the subjects failed to sustain

[12] Appendix Table B.6 reports a regression analysis to explain subjects' decisions to contribute as a function of the sanction rates enforced. The analysis found that the size of the sanction rate enacted in a group was a significantly positive predictor of the members' contribution amounts.

cooperation when non-deterrent sanction rates were instead enacted (again, see Table 4 and Panel A of Figure 4). While this is also expected, since free riding is clearly the unique Nash Equilibrium outcome with mild sanctions, the contribution levels in the Voting-D treatment were persistently extremely low (the connected dotted lines in Figure 4). This was not the case for the Voting-N treatment, although non-deterrent cases were observed only in Phases 2 and 5 here. The extremely low contribution level with non-deterrent FS suggests that depleted subjects in Voting-D could not resist the temptation to free ride with only mild law. On average, approximately 70% of the subjects voted for deterrent sanction rates in both the Voting-D and Voting-N treatments (Panel C of Figure 4).

One interesting difference was observed between the states of self-regulatory resources: the average contribution under the IS scheme exhibits a significantly increasing trend in the Voting-D treatment. This is in contrast to the trend in the Voting-N treatment, where it remains stable or is somewhat in a decreasing trend (Panel A of Figure 4, Part II of Appendix Table B.4). The same is observed for the subjects' payoffs (Panel B of Figure 4). This difference in the efficiency trend can be explained by the significantly strong pro-social punishment of the depleted subjects in the Voting-D treatment relative to that in the Voting-N treatment (Panel D of Figure 4, Appendix Table B.8). Here, 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 $\overline{c_t}$ in Period *t*. While pro-social punishment was found to be significantly stronger than anti-social punishment in both the Voting-D and Voting-N treatments (Appendix Table B.9), the punishment activities among the *depleted* subjects were more intense, much more than double those among the non-depleted subjects (Panel D of Figure 4).[13] This feature matters in fostering cooperation norms because those who had been pro-socially punished in Period *t* increased contribution amounts in Period $t + 1$, *ceteris paribus* (Appendix Table B.10).

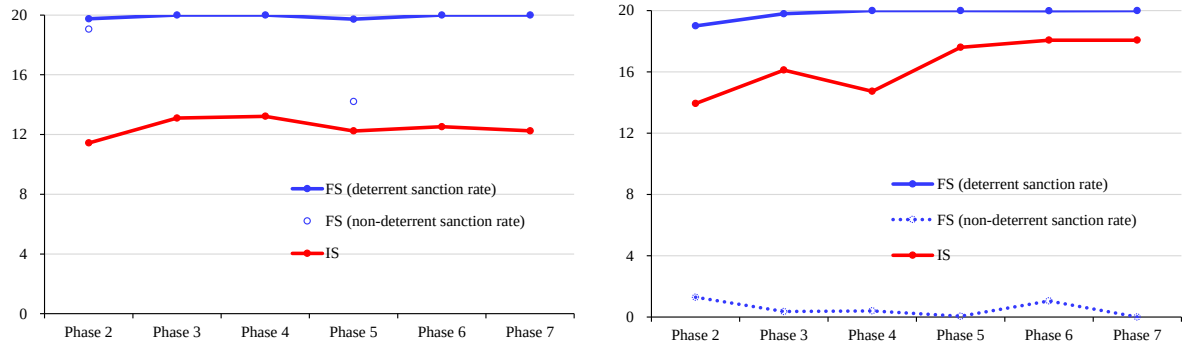
Note that subjects have two conflicting sources of temptation in the punishment stage under the IS scheme: one is to free ride on peers' punishment acts, while the other is to inflict justice driven by negative emotions. The observed punishment patterns suggest that depleted individuals succumbed to the *hot* temptation to respond to their negative emotions, rather than to their *cool* temptation to free ride on others' punishment acts. This interpretation supports 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. There are, of course, many 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 still preferred to rely on the FS scheme even after gaining experience. The welfare loss due to punishment activities and administrative cost payments was much smaller in the IS scheme than in the FS scheme in a later phase—see Appendix Figure B.2.

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

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

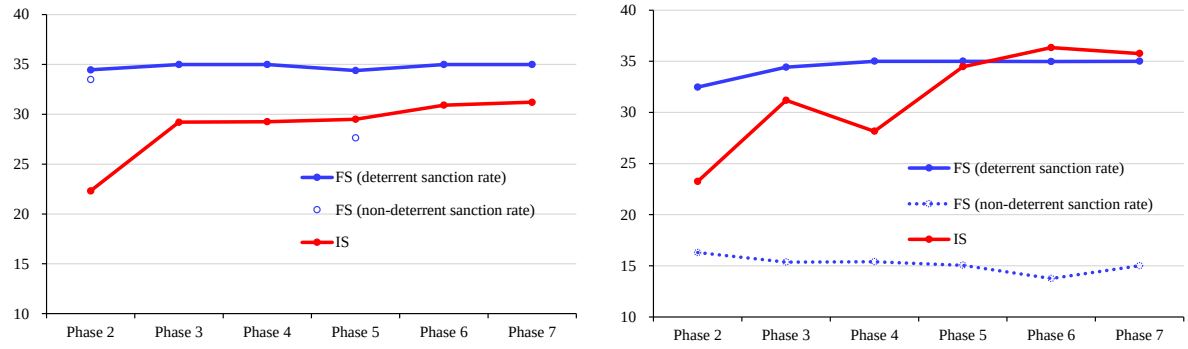
punishment was stronger than anti-social punishment for both voting treatments. Interestingly, pro-social punishment by depleted subjects was significantly stronger than that by non-depleted subjects, and thus depleted subjects more effectively strengthened cooperation norms over time in the IS scheme.

Figure 4: *Performances under Selected Schemes*



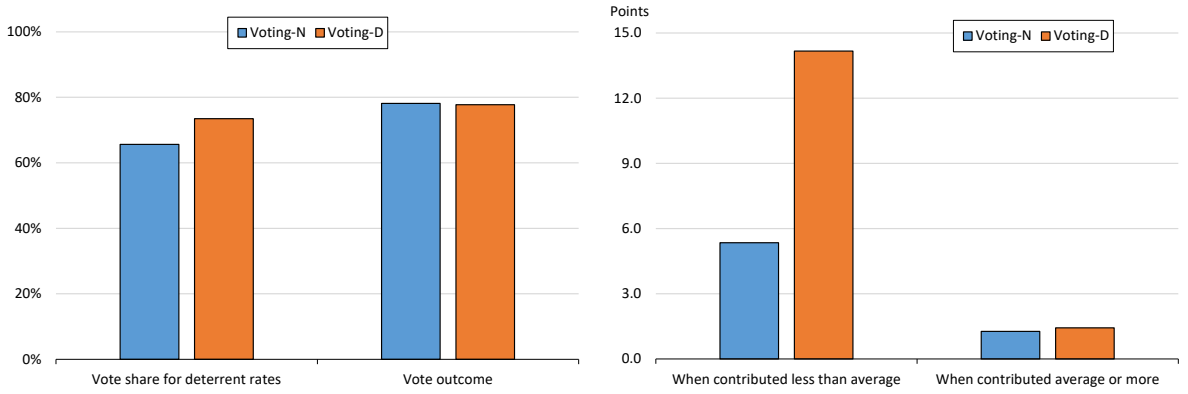
i. Voting-N treatment[#1] ii. Voting-D treatment

A. Average Contribution by Sanction Scheme



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 FS D. Avg. per Period Loss due to Punishment Received under IS[#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 4: *Deterrence of the FS Scheme and Efficiency*

Dependent variable:	Contribution of Subject *i* in Period *t*, where *t* > 4				Payoff of Subject *i* in Period *t*, where *t* > 4			
Independent variables:	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.

6. Supplementary Opinion Survey

While the laboratory experiment revealed strong effects of people's self-regulatory resources on their policy preferences, one concern is its external validity, as is sometimes the case for a neutrally framed laboratory experiment. To supplement the main laboratory experiment, an opinion survey was additionally conducted regarding people's self-control behaviors and their policy preferences during the recent Covid-19 pandemic. The survey was conducted in July 2022 by recruiting third- and fourth-year undergraduate students at Kansai University.[14] As explained below, it was found that those with weaker self-control were more likely to support the strengthening of the formal enforcement of self-restraint behavior, consistent with the main result from the laboratory experiment.

A challenge in collecting the information of self-control behaviors from respondents is the presence of a possible social desirability bias. Considerable prior experimental research has shown that people are reluctant to accept their socially undesirable behaviors when directly asked in a survey. For

[14] 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. Note that while Japan has declared a state of emergency four times thus far, third- or fourth-year undergraduates experienced all the four self-restraint requests as university students.

example, in the context of an election, respondents are reluctant to accept their vote-buying experiences (e.g., Gonzalez-Ocantos *et al.*, 2012). To avoid such a bias, the respondents were provided with ten concrete examples in the survey, among which seven were on weak self-control behaviors (e.g., "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 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 for the detail). There are two benefits of using the approach just stated: First, it is possible to let respondents consider more concrete behaviors than when a question asks about their self-control in an abstract manner (e.g., did you comply with almost all the restriction measures imposed in the region?), thereby making it possible to have a more precise measure of their self-control. Second, the ten examples include both socially desirable and undesirable behaviors, whose aspects make 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, the respondents' answers were expected to range from 3 to 7, such that a larger number would correspond to a weaker self-control type.[15]

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 subjects' perceptions of others' self-control (Appendix C.1.2). Not only people's preferences for strong formal restrictions and penalty, but also their own self-control behaviors may be affected by their beliefs about others' behaviors, in which case an omitted variable bias may influence the result. Note that conditional cooperation is quite common in the context of a social dilemma (e.g., Fischbacher *et al.*, 2001; Fischbacher and Gächter, 2010). The questionnaire also asked questions on a variety of respondents' demographic and background information as control variables. These questions are included in Appendix C.1.4.

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 the strengthening of the formal enforcement of restrictions. This significant correlation is not affected by whether people's perceptions of others' self-control or any other control variable are added. The result is also robust to the regression method used—a linear or tobit regression. The supplementary survey therefore confirms, in a realistic context, the key result of a significant relationship between people's self-control types and their

[15] 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.

commitment behaviors that was obtained in the main laboratory experiment.

7. Conclusion

Social dilemmas are ubiquitous in both our private and economic lives, while people's free riding in such dilemmas is known to be harmful to societies and organizations. During the last few decades, economic research has documented that the dilemmas can be overcome when people's incentives are changed by enforcing a formal institution. 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 unsettled 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). The present 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 rigorously manipulated their self-regulatory resources, most of the 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. A survey on the Covid-19 pandemic revealed a similar relationship: those who had weaker self-control attitudes were more likely to support stronger restriction measures.

This study is closely related to a large research area on self-control and self-regulatory resources. The scheme choice preference found in the experiment is consistent with the well-known self-control model formalized by Gul and Pesendorfer (2001, 2004) when combined with social preferences. The theory suggests that people with *small* amounts of self-regulatory resources incur a *large* self-control cost when they exercise self-control, such that they do not succumb to the temptation to free ride. The presence of disutility causes them to remove such temptations by voting in advance as a commitment device. In contrast, people can easily resist temptation when their resources are abundant. Thus, such strong self-control types sustain cooperation with informal punishment, rather than enact costly formal punishment. The present experiment underlines the presence of such human self-control preferences and the predictive power of the commitment theory in the context of endogenous institutional formation.

While the result obtained from the experiment is sufficiently clear, it is worth emphasizing that the present study is only the first step in exploring the role of self-regulatory resources in an institutional setting. For example, this study adopted experimental parameters frequently used in the literature, such as the group size of five, MPCR of 0.4, and 24-period interactions in Part 2. These settings are desirable and standard, and satisfy the usual requirements for a fixed laboratory size and an experiment duration of approximately two hours. However, there are numerous 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 with different game parameters. For another example, the accuracy of enforcement and noise may affect people's institutional formation. The present study assumes, for simplicity, that not only do the subjects accurately observe their peers' contributions, but that the 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 a real authority or society. A novel experiment by Nicklisch *et al.* (2016) demonstrated that the imperfect observability of peers' contributions (hence some noise in punishment) raised the attractiveness of formal mechanisms when anti-social decentralized punishment was severe in a group. It can be imagined that self-regulatory resources may be more important for people to behave in such a complex, risky environment; however, how the resource amount affects their institutional choices is unclear. Further experimental research would certainly be useful before the role of self-control is generalized in the context of institutions and social dilemmas.

References

Achtziger, Anja, Carlos Alós-Ferrer, and Alexander K. Wagner, 2016. The Impact of Self-Control Depletion on Social Preferences in the Ultimatum Game. *Journal Economic Psychology*, 53, 1-16.

Anderson, Christopher, and Louis Putterman, 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, James, Laura Gee, 2012. Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, 96(11-12), 1036-1046.

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

Baumeister, Roy, Todd Heatherton, and Dianne Tice, 1994. *Losing Control: How And Why People Fail At Self-regulation*. San Diego, CA: Academic Press, Inc.

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

Bednar, Jenna, Yan Chen, Tracy Liu, and Scott Page, 2012. Behavioral spillovers and cognitive load in multiple games: An experimental study. *Games and Economic Behavior*, 74(1), 12-31.

Bucciola, Alessandro, Daniel Houser, Marco Piovesan, 2011. Temptation and Productivity: A Field Experiment with Children. *Journal of Economic Behavior & Organization*, 78, 126-136.

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

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

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

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

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

Dal Bó, Pedro, Andrew Foster, and Kenju Kamei. 2019. The Democracy Effect: a Weights-Based Identification Strategy. NEBR Working Paper 25724.

de Quervain, D.J., Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Urich Schnyder, Alfred Buck, and Ernst Fehr, 2004. The Neural Basis of Altruistic Punishment. *Science*, 305(5688), 1254-1258.

Eddie Dekel, Barton L. Lipman, Aldo Rustichini, 2009. Temptation-Driven Preferences. *Review of Economic Studies*, 76(3), 937-971.

DeWall, Nathan, Nicole Mead, Roy Baumeister, and Kathleen Vohs, 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, Arhan, Talbot Page, and Louis Putterman, 2009. Who to punish? Individual decisions and majority rule in mitigating the free rider problem. *European Economic Review* 53(5), 495-511.

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

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

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

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

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

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

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

Fehr, Ernst, and Klaus Schmidt, 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, Ernst, and Tony Williams, 2018. Social Norms, Endogenous Sorting and the Culture of Cooperation. IZA Discussion Papers 11457.

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

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

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

Gächter, Simon, Elke Renner, and Martin Sefton, 2008. The long-run benefits of punishment. *Science*, 322(5907), 1510-1510.

Gerhardt, Holger, Hannah Schildberg-Hörisch, and Jana Willrodt, 2017. Does Self-Control Depletion Affect Risk Attitudes? *European Economic Review*, 100, 463-487.

Gonzalez-Ocantos, Ezequiel, Chad de Jonge, Carlos Meléndez, Javier Osorio, David Nickerson, 2012. Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua, *American Journal of Political Science*, 56, 202-217.

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

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

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

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

Gürerk, Özgür, Bernd Irlenbusch, and Bettina Rockenbach, The competitive advantage of sanctioning institutions. *Science*, 312 (5770), 108-111.

Hauser, Oliver, David Rand, Alex Peysakhovich, and Martin Nowak, 2014. Cooperating with the future. *Nature*, 511, 220-223.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008, Antisocial punishment across societies. *Science*, 319 (5868), 1362-1367.

Houser, Daniel, Daniel Schunk, Joachim Winter, and Erte Xiao, 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, Kenju, 2011. Self-Regulatory Strength and Dynamic Optimal Purchase. *Economics Letters*, 115(3), 452-454.

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

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

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

Kamei, Kenju, 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, Kenju, Louis Putterman, and Jean-Robert Tyran, 2015. State or Nature? Endogenous Formal versus Informal Sanctions in the Voluntary Provision of Public Goods. *Experimental Economics*, 18, 38-65.

Kamei, Kenju, and Katy Tabero, 2021. The Individual-Team Discontinuity Effect on Institutional Choices. Working paper.

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

Kocher, Martin, Peter Martinsson, Kristian Myrseth, Conny Wollbrant, 2017. Strong, bold, and kind: self-control and cooperation in social dilemmas. *Experimental Economics*, 20, 44-69.

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

Hoxby, Caroline, 2000. Does Competition Among Public Schools Benefit Students and Taxpayers? *American Economic Review* 90(5), 1209-1238.

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

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

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

Masclet, David, Charles Noussair, Steven Tucker, and Marie-Claire Villeval. 2003. Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism. *American Economic Review*, 93(1), 366-380.

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

Mead, Nicole, Roy Baumeister, Francesca Gino, Maurice Schweitzer, Dan Ariely, 2009. Too Tired to Tell the Truth: Self-Control Resource Depletion and Dishonesty. *Journal of Experimental Social Psychology*, 45(3), 594-597.

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

Nicklisch, Andreas, Kristoffel Grechenig, and Christian Thöni, 2016. Information-sensitive Leviathans. *Journal of Public Economics*, 144, 1-13.

Nikiforakis, Nikos, and Hans-Theo Normann, 2008. A comparative statics analysis of punishment in public-good experiments. *Experimental Economics*, 11, 358-369.

Putterman, Louis, Jean-Robert Tyran, and Kenju Kamei, 2011. Public goods and voting on formal sanction schemes. *Journal of Public Economics*, 95(9-10), 1213-1222.

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

Stroop, Ridley, 1992. Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology: General*, 121(1), 15-23.

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

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

Traulsen, Arne, Torsten Röhl, and Manfred Milinski, 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, Jean-Robert, and Lars Feld, 2006. Achieving Compliance when Legal Sanctions are Non-deterrent. *Scandinavian Journal of Economics*, 108(1), 135-156.

Zhang, Boyu, Cong Li, Hannelore De Silva, Peter Bednarik, and Karl Sigmund, 2014. The Evolution of Sanctioning Institutions: An Experimental Approach to the Social Contract. *Experimental Economics*, 51(2), 285-303.