

## TITLE: INTRINSIC DISUTILITY OF PUNISHMENT: A SURVEY EXPERIMENT

*Peter Lewisch (University of Vienna)*

*Stefania Ottone (University of Turin)*

*Ferruccio Ponzano (University of Eastern Piedmont)*

### LONG ABSTRACT

#### PROJECT DESCRIPTION, RESEARCH QUESTION, HYPOTHESES

The social sciences, and economics in particular, have emphasized for a long time the crucial role of punishment as a tool to achieve social stability. Punishment, however, is costly for the punisher. Usually, interdisciplinary studies underline the private cost of punishment (Galle and Mungan 2021) as well as both its social and private benefits (Levy, 2022). In an instrumental perspective, punishment is likely to be carried out, if expected individual benefits are higher than individual costs. Benefits of punishment for the more encompassing group do not enter the individual cost-benefit calculus, leading in many instances to an underenforcement of punishment. ‘Altruistic punishment’ (i.e., costly punishment that serves no instrumental goal for the punisher) could serve, as suggested by the pertinent experimental literature, as a powerful enforcer of social norms. The category of “altruistic punishment” covers all non-instrumental motivations, ranging from considerations of just desert and justice to the “sweet taste of revenge”. The possibility of such “utility gains” for the punisher increase the individual benefits of punishment and incidence of its occurrence. We intend to examine experimentally the existence and impact of the “inverse category”, namely a possible negative utility that people may experience when punishing another person. To our knowledge, no economics study exists that focuses on intrinsic disutility of punishment. The basic idea of this “intrinsic disutility of punishment” has been formulated by nobel prize laureate James Buchanan (Buchanan 1975, 130), describing a second category of punishment costs that goes beyond the ‘regular’ resource components (for investigations, proceedings, etc.) required for penal enforcement. This second category captures the negative emotions by the punisher associated with the deliberate infliction of a bad onto someone else. ‘Intrinsic disutility’ accounts for the straightforward fact that people normally do not like to harm another being. As Buchanan (1975, 133) states:

The basic costs of punishment are subjective, and these can best be conceived in a utility dimension.

The imposition of penalties on living beings, whether or not these beings have violated law, causes pain, utility loss, to the normal person who must, directly or indirectly, choose these penalties.

‘Punishing others’ is a ‘bad’ in economic terms, an activity that is, in itself undesirable, an activity that normal persons will escape if possible or, failing this, will pay to reduce.

Whereas this basic idea appears to have an immediate explanatory appeal, it has not been followed up in the literature. However, there have been few observations in the literature (in particular, Carlsmith et al. [2008], *The Paradoxical Consequences of Revenge*) that, albeit without any reference to the concept of an 'intrinsic disutility of punishment' report the "surprising" result of not having found the expected relief when they analyzed people's satisfaction after punishing. Outside any punishment context, the current psychological empathy research has also found that people would be willing to spend resources to reduce the suffering of unrelated third parties, suggesting a non-instrumental psychological interest in stopping harm to others, even if this comes at a personal cost. This implies that, while intrinsic utility of punishment increases penalty, intrinsic disutility reduces punishment. This may have an impact on stability of cooperation. Forgiveness, from a moral point of view, is a "good thing"; perhaps not so good, if strategically exploited by wrongdoer relying on intrinsic disutility of punishment.

Our research question is: in a third-party (altruistic) punishment scenario, which factor (the intrinsic utility or the intrinsic disutility) will dominate under what conditions?

We focus on third-party punishment since it has been defined as the Golden Keystone of Human Cooperation and Social Stability (Lewisch, 2020 for a brief survey). In our study, we want to detect whether personal traits (like prosociality and empathy) and the severity of a crime (assault vs theft) may influence the relative weight of intrinsic disutility of punishment. We will test the intrinsic disutility of punishment detecting subjects' propensity to avoid punishing by themselves.

Our hypotheses are based on the reasonable idea that which factor will dominate is likely to depend both on framing of relevant choice setting and on personal characteristics. In particular:

H1: Intrinsic disutility of punishment is higher when the established sanction is physical

H2: Intrinsic disutility of punishment is lower when the crime involves the physical dimension

H3: Intrinsic disutility of punishment is higher for prosocial and empathic people

H4: Intrinsic disutility of punishment is lower in countries where sanctions are harsher

## **METHODOLOGY: EXPERIMENTAL DESIGN (DRAFT) AND SAMPLE SIZE**

We will implement a Survey on Qualtrics in Italy and Germany. The choice of the two countries depends on the fact that in Germany there are sanctions that involve the physical dimension (such as chemical castration as a voluntary measure to reduce time of imprisonment), while in Italy it is not the case. Moreover, in Germany theft is punished with more than double the number of years in prison compared to Italy. This implies that in Germany people are used to harsher sanctions. The

experimental design is based on a variation both in the crime (assault and theft) and in the nature of the sanction (monetary or physical). Thus, 4 treatments are created (assault-monetary sanction; assault-physical sanction; theft-monetary sanction; theft-physical sanction). In each treatment, participants are informed about the crime and the established sanction. At this point, they are asked to choose whether they want to punish the guilty subject personally or not.

In each country the questionnaire will be administered to 1000 resident citizens (500 women and 500 men) aged between 18 and 65. In each country two experimental groups will be created: one group will participate in treatments where the crime is assault while the other group will make decisions for theft. Each group will be asked to make decisions both when monetary sanctions are involved and when the sanctions are physical. The order in which the treatments are presented will be random.

At the end of each session, we collect data on participants' personal characteristics through a brief socio-demographic questionnaire, SVO and IRI. Subjects' satisfaction is detected too.

The survey will last 25 minutes.

The sample size guarantees a balanced number of observations between countries and genres. Moreover, it allows to run parametric tests and robust econometric analyses with different sets of explanatory variables.

## **PROJECT TIMELINE AND EXPECTED RESULTS**

We we collect data by the end of November. This will allow to present a preliminary analysis at the SIDE conference.

Our expected results are related to the four hypotheses presented in the previous section. In particular, we expect that intrinsic disutility of punishment is affected by four dimensions: the type of sanction, the type of crime, the type of person and the country where people live. Our study aims at detecting the relative importance of each of those variables.

## **SELECTED REFERENCES**

- Buchanan, J. (1975), *The Limits of Liberty: Between Anarchy and Leviathan*, Chicago
- Carlsmith, K. M., Wilson, T. D., and Gilbert, D. T. (2008). "The paradoxical consequences of revenge". *Journal of Personality and Social Psychology*, 95 (6)
- Galle B., and Mungan M. C. (2021), "Optimal Enforcement with Heterogeneous Private Cost of Punishment", *Journal of Legal Studies*, 50 (2).
- Levy D. A. (2022), "Optimizing the social utility of judicial punishment: An evolutionary biology and neuroscience perspective", *Frontiers in Human Neuroscience*, 16.
- Lewis, P. (2020), "Altruistic Punishment: The Golden Keystone of Human Cooperation and Social Stability?", *Analyse und Kritik*, 42 (2).