

The Damocles effect:

Judges may inflate the duration of suspended prison terms by over 50%

Andrzej Uhl^{1,2*}

¹ University of Cambridge, United Kingdom

² St John's College, Cambridge, United Kingdom

* Corresponding author. Email: au274@cam.ac.uk

Manuscript accepted for publication

in the Journal of Experimental Criminology on 3 March 2025

Acknowledgement: I gratefully thank Piotr Girdwoyń for granting access to the data and his insights into the sentencing practice. I am also indebted to Paweł Ostaszewski for sharing criminal justice statistics on suspended sentences. Credit is due to Monika Gniadek-Wąż and four anonymous reviewers for their helpful comments.

Conflict of interest: The author declares that he has no conflict of interest.

Abstract

Objective: Legal scholars suspect that judges choose longer prison terms when they are going to suspend the sentence. This study examines this so-called sentence inflation in a controlled condition, holding case-related confounds constant.

Methods: I analyse the differences between suspended and unsuspended prison terms in the data from the Polish judicial exam. Each judge ($N = 232$) sentenced the same case based on a detailed court file. Judges had high stakes in the exam and spent over six hours choosing and justifying the sentence.

Results: Many judges sentenced the offenders to prison. The suspended prison terms were 60% to 168% longer than unsuspended prison terms meted out in the identical case.

Conclusions: Judges display a tendency to inflate suspended prison terms, perhaps in order to appease the punitive public and strengthen individual deterrence. With high reoffending rates, this well-intentioned practice might backfire, leading to a surge in the prison population.

Key words: sentencing, suspended sentence, sentence inflation, judicial decision-making, experimental jurisprudence

Introduction

Do defendants receive longer prison terms if the judge chooses to suspend the sentence? They should not, if only because the law makes no allowance for such distinctions. The choice of punishment and the decision to suspend it are separate (Ashworth, 1983; Morris & Tonry, 1991). By choosing the punishment, the judge decides what sentence fits the offenders' crime. By suspending the prison sentence, the judge determines that the defendant will not be immediately imprisoned, provided he or she abides by set conditions. If the probationer violates these conditions during the probation period, the judge can (or, in some cases, must) revoke the suspension and order immediate imprisonment. However, if the probation period ends without revocation, the offender may never set foot in a prison. Such a second chance is often granted when the defendant has no prior criminal record and the risk of reoffending appears low. While certain offender characteristics may influence both decisions (sentence duration and suspension), the decisions themselves should not influence each other. Judges should not choose to suspend a sentence based on its duration, nor should they make the sentence longer simply because it is being suspended. In other words, the length of the sentence and the decision to suspend it ought to be independent.

However, scholars suspect that judicial decisions deviate from a purely mechanistic application of binding law (e.g., Rachlinski & Wistrich, 2017; Spamann & Klöhn, 2024). Writing for *Crime and Justice*, Krajewski (2016) speculates that judges in Poland make suspended sentences longer in order to enhance deterrence in the absence of immediate punishment and to counteract public perceptions of impunity associated with such sentences. The issue is not purely legalistic. Given limited support for probationers and high reoffending rates, this practice may prove harmful in the long run, creating a dead weight of prisoners serving disproportionately long sentences for minor but repeated offenses. Krajewski's argument is worth quoting at length:

“judges may foresee that they will suspend a sentence and only later on decide on its length, making it longer than an executed sentence in a similar case (...) to reinforce an offender's motivation not to reoffend (...) [and] for the sake of public perceptions: longer imprisonment is supposed to compensate somehow for the suspension. This practice may have serious consequences. Offenders whose suspended sentences are revoked may spend more time in prison than do those originally sentenced to incarceration.”

The practice of increasing the length of suspended sentences has also raised concerns among legal scholars in other jurisdictions. Bottoms (1981) suggested that after suspended sentences became available in England, “magistrates’ courts began to impose longer sentences in cases where they suspended the sentence than in cases where they actually imposed imprisonment” (p. 6). A similar pattern may have occurred in Australia, where the term “sentence inflation” was coined to describe the additional prison time given to offenders whose sentences were suspended (Tait, 1995; Bagaric, 1999). These scholars’ primary concern was that when some of these inflated sentences become activated—that is, when the suspension is revoked—the prison population may increase, despite the policy’s original intent to keep offenders out of prison.

How can we determine whether judges inflate the duration of suspended sentences? The comparison of real-life sentences suffers from a fundamental problem of causal inference (Holland, 1986): once a judge has chosen to suspend a sentence, we cannot know what its duration would have been had he or she opted for an unsuspended sentence. Researchers cannot make a valid comparison between suspended and unsuspended sentences based on observational data, since cases are rarely equivalent. More serious offenses, which warrant longer prison terms, are often committed by serious offenders, who are less likely to receive a suspended sentence. Additionally, the law may restrict the suspension of longer sentences. For instance, in Poland, when the current data was collected, the maximum “suspendable” sentence was three years. This may explain why even scholars concerned about sentence inflation have sometimes presented statistics showing that unsuspended prison terms tend to be longer than suspended ones (Bartels, 2008; Krajewski, 2016). The key question is whether an individual offender is likely to receive a longer sentence simply because it is suspended.

I address this question using a unique dataset: sentences passed by trainee judges as part of their judicial exam. Each judge was given an identical court file of over 100 pages and six hours to compose a complete judgment, including the sentence. The judges, who were permanently appointed a few months later, had a high stake in the exam, incentivizing them to carefully consider the case. This approach advances experimental jurisprudence (see Sommers, 2021; Holste & Spamann, 2023) and improves upon the commonly used short sentencing vignettes, which judges often complete with no skin in the game—a method criticized elsewhere (Fischman, 2013; Pina-Sanchez & Linacre, 2013,

2016). The procedure closely approximates an (otherwise infeasible) “experiment”¹ in which several hundred judges independently sentence the same case, unaware of one another’s decisions. It is also well suited to the research question at hand; since judges often opted for either a suspended or an unsuspended prison sentence, we can directly compare the durations of these sentences. This comparison eliminates potential case-related confounds, as the cases assigned suspended and unsuspended sentences were identical.

Legal Context

The institution of suspended prison sentences varies across jurisdictions and may change dynamically within each jurisdiction. Relevant to this research is Polish criminal law as of March 2023 (Polish Penal Code, Articles 69–76). Below, I briefly outline the general rules that applied to most cases at the time, including those presented in the judicial exam.

Judges do not follow sentencing guidelines but instead select punishments from broad sentencing ranges specified for each type of offense. Except for the most serious crimes, these ranges allow judges to impose either a fine, community service, or a prison term capped at a certain number of months. Of these sentencing options, only prison terms can be suspended. Suspension is never mandatory and is only available if three cumulative conditions are met:

1. The prison term cannot exceed one year.²
2. The offender must not have been serving a prison sentence for another crime at the time of the offense.
3. A suspended sentence must be sufficient to achieve the objectives of punishment, particularly the prevention of reoffending.

The judge may attach specific conditions to the suspension. The judge also determines the probation period, which typically differs from the duration of the suspended prison term. If the offender (now a probationer) violates the conditions, the judge may revoke the suspension and order immediate imprisonment. Revocation is mandatory if, during the probation period, the probationer

¹ The current procedure is not an experiment in the narrow sense, since it lacks manipulation (all judges were subject to the same treatment condition).

² Note what this implies: judges should know the exact duration of the prison term by the time they consider whether to suspend the sentence.

commits a deliberate offense similar to the original offense. Other breaches of the conditions (e.g., failure to meet with a probation officer) may also be reported to the court and are subject to judicial discretion. If the probation period passes without revocation, the sentence is automatically expunged.

Methods

Participants

A total of 271 candidates sat the exam, including 196 trainee judges and 75 individuals holding lower-tier judicial positions, such as judicial assistants or court referendaries. Access to judicial training was restricted to graduates of a five-year law degree who had passed a highly selective entry exam (<10% admission rate). The intensive three-year training program included supervised judicial work in regional courts across the country. Candidates were allowed only two attempts in their lifetime, and the sample included some retakers ($N = 17$). Candidates who failed the written part of the exam were excluded from the analysis, yielding an analytic sample of 232. These remaining candidates then attempted the oral part, leading to further attrition of 44 candidates. Since they were given new identifiers in the oral part, these failed candidates could not be excluded from the dataset. However, achieving a passing mark in the sentencing tasks suggests that their criminal sentencing skills were deemed comparable to those of the rest of the cohort, with their shortcomings lying in other areas of law. The successful candidates ($N = 188$) were appointed to permanent judicial positions a few months after the exam, and most currently serve as active judges in regional courts nationwide.

Materials

The simulated court file contained a generic criminal case that early-career judges might encounter in their work. Its 104 pages included all documents typically found in a court file, beginning with a police report and continuing through the indictment, evidence, and witness testimonies, culminating in the final pleadings by the parties to the trial. Relevant to this research were the actions of two offenders. A 38-year-old man broke into a local liquor store and stole three bottles of whisky with a market value of about 360 PLN (approx. 90 USD), constituting *burglary*. Shortly thereafter, another 27-year old man entered the already burglarized shop and took two bottles of vodka worth 240 PLN (approx. 60 USD), constituting *theft*. Both offenders were arrested, briefly

held in custody, and then released pending trial. Neither had a prior criminal record. The prosecutor sought a one-year suspended prison sentence for both men.

Procedure

The exam was centrally administered by the Ministry of Justice at a single location on 7 March 2023. Judges sat the exam under supervised conditions, with no communication allowed. They composed their judgments on a computer, with access restricted to a text editor and the court file document. They were also permitted to bring a paper copy of the Penal Code and related legislation. Candidates had 390 minutes to draft a full judgment, including its justification. A typical judgment consisted of dispositive rulings concerning the offender, followed by a detailed analysis of the evidence, the legal basis, and the reasoning behind the chosen punishment. After the judgments were submitted and assessed by examiners, the National School of the Judiciary and Public Prosecution took custody of the records for archival purposes.

Measures

Following the release of files, courtesy of the National School of the Judiciary and Public Prosecution, I recorded each judgment as an individual observation, starting with the exam score on a 0–60 scale. The values in the analytic sample are left-censored at the lowest passing mark (17). The specific dispositive rulings were then extracted from the judgment files.

At the time of sentencing, the law permitted various sentencing options in this case, including diversion from prosecution, a fine, community service, or a prison term. The following analysis focuses on the subsamples of judges who imposed a prison sentence for burglary and/or theft ($N = 148$, 62.1%; $N = 20$, 8.6%, respectively). Sentences other than imprisonment (e.g., fines) could not be suspended and are therefore irrelevant to the analysis of sentence inflation (see above). Prison sentences were measured in months, with a statutory maximum of 60 months for both offenses under consideration. However, only sentences of up to 36 months were eligible for suspension. In practice, all sentences imposed were 12 months or shorter, meaning that every unsuspended sentence in the sample could have been suspended. This is a key point, as it ensures that “non-suspendable” sentences above a certain threshold do not distort the analysis. Suspension of the sentence was treated as a binary variable.

Findings

Sentences for burglary

Of the 148 judges who sentenced the first offender to a prison term, 128 imposed a suspended prison sentence, while 20 imposed an unsuspended prison term. Figure 1 presents the distribution of suspended and unsuspended prison sentences. The mean duration of unsuspended prison sentences was 4.85 months (median = 4, $SD = 2.18$). In contrast, suspended prison sentences were significantly longer, with a mean of 8.11 months (median = 6, $SD = 3.57$). The difference in sentence duration between suspended and unsuspended prison terms was substantial (Cohen's $d = 0.95$, indicating a large effect size) and statistically significant according to a Mann-Whitney U test ($W = 631, p < .001$).

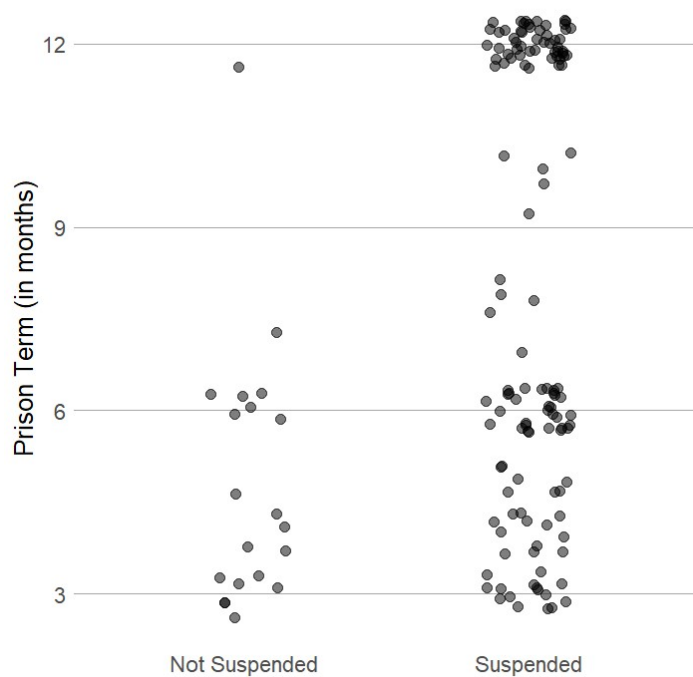


Figure 1. Jitter plot of sentence duration by suspension status.

Next, I examine whether sentence inflation was moderated by judges' general sentencing skills, approximated by their exam scores.³ High-scoring judges might have a better grasp of sentencing law and, consequently, make independent decisions regarding sentence duration and suspension. This hypothesis was tested using an OLS regression model with robust standard errors, including a dummy

³ Naturally, judges could not, and did not lose points for inflating sentences, since sentence inflation can only be detected in aggregates.

variable for sentence suspension, exam score, and their interaction (product term). The coefficient for the interaction term was not statistically significant ($b = 0.12$, $SE = 0.07$, $p = .103$), suggesting that even the highest-rated judges did not resist the tendency to inflate suspended prison terms.

Alternatively, however, this null finding could be due to statistical power insufficient to detect a moderation effect.

Sentences for theft

Far fewer judges ($N = 20$) imposed a prison sentence on the second offender. Two sentenced the offender to immediate imprisonment, while 18 imposed a suspended sentence. The mean duration of unsuspended prison sentences was 2.5 months (median = 2.5, $SD = 0.71$), whereas the mean duration of suspended prison sentences was 6.67 months (median = 6, $SD = 3.08$)—a 266% increase compared to the mean unsuspended sentence. The difference was statistically significant in a Mann-Whitney U test ($W = 1$, $p = .035$). This finding suggests that sentence inflation may be particularly pronounced at the lower end of the sentencing range, where judges can more than double the sentence length if they choose to suspend the sentence, yet the sample size limits the reliability of this finding.

Discussion

The probationer's situation has often been likened to that of Damocles—an ancient courtier placed by his king beneath a sword suspended by a single horsehair (Bottoms, 1981; Bartels, 2009). Any careless movement could bring disastrous consequences—serving as a “clear warning to the offender, which he disregards at his peril” (Ashworth, 1983, p. 437, sic). This report examined whether judges tend to “sharpen” the sword of Damocles (Braithwaite, 2018) by imposing longer prison terms when they intend to suspend them. To answer this question with confidence, I analyzed data from a judicial exam in Poland, in which trainee judges sentenced an identical case under high-stakes, carefully controlled conditions. Even though all case facts were held constant, suspended prison terms for burglary were, on average, 59.8% longer than unsuspended prison terms. This finding aligns with the 50% inflation rate reported by Tait (1995) in observational data from Australian magistrates. In the smaller sample of sentences for theft, the inflation rate was even higher at 168%. These results suggest that Polish judges determine the duration of prison terms based, at least in part, on their—presumably prior—decision to suspend the sentence.

Why would judges deviate from the letter of the law and inflate suspended sentences? One possible motivation is to appease the punitive public, which may perceive suspended sentences as a mere slap on the wrist (Freiberg & Moore, 2009; Bartels, 2010; van Gelder et al., 2015). Inflating the length of a suspended sentence might serve as compensation for its suspension, seemingly at no immediate cost. Judges may also have another, related rationale for inflating suspended sentences: a longer suspended sentence is an intuitively comprehensible appeal to a reasoning offender (Wasik, 1994). While it allows the offender to avoid the immediate pain of incarceration (Bottoms, 1981), it simultaneously increases their stake in conformity—should they reoffend, the long suspended sentence would be activated and added to the punishment for the new offense.

Both of these motivations betray undue optimism. Many convicted offenders do recidivate and individual deterrence through punishment severity is at best a modestly effective instrument of behavioural change (Dölling et al., 2009; Doob & Webster, 2003; Nagin, 2013). This means that in many hypothetical worlds, the fictitious burglar from the exam file would eventually re-offend and face lengthy, perhaps disproportionately lengthy prison time. Recidivism is still more likely in an inefficient criminal justice system where “many offenders during the suspended sentence period are left alone with their alcohol, drug, mental health, and other problems,” and—especially if supervised by overburdened probation officers —“persist in bad habits, reoffend, or otherwise attract attention that lands them in prison” (Krajewski, 2016, p. 214). A broader implication is that well-intentioned attempts to keep offenders out of prison can easily backfire, and produce a dead weight of prisoners serving long sentences for surprisingly light, if repeated, offences.

In fact, Krajewski (2016) suggests that the inflated suspended sentences might be among the reasons why Poland has the highest prison rate in the European Union, keeping its second largest prison population (70,418 prisoners, that is 193 per 100,000 inhabitants; World Prison Brief, 2024). Polish sentencing policy, he writes, found itself in a catch-22 of sorts; efforts to reduce incarceration through increased use of suspended sentences may, paradoxically, drive up already high imprisonment rates (Krajewski, 2016, 2023). Consider that in 2010s Poland, almost a quarter of suspended prison sentences were revoked (Nawój-Śleszyński, 2014), and more prisoners served time following revocation than as a result of immediate imprisonment (Mycka & Kozłowski, 2013). Paweł

Ostaszewski of Poland's Institute of Justice has calculated that, in 2024, courts revoked 19.8% of sentences suspended in that year and 18.1% of sentences suspended in the previous year.⁴ All this suggests that the inflation of suspended prison sentences—a seemingly minor deviation from the legalistic sentencing model—could have a disproportionate impact on the prison population, as the probationers-turned-prisoners end up spending much more time behind bars.

Limitations

Two key limitations should be noted. First, while the procedure held all case-related factors constant, it did not account for person-related confounders. One unmeasured variable likely influencing both sentence length and the decision to suspend a sentence is the punitiveness of individual judges. However, more punitive judges should impose longer sentences and be less likely to suspend them. This suggests that a suppression effect is more plausible than a spurious correlation—meaning the true sentence inflation rate could be even higher if judicial punitiveness had been controlled for. Second, the analysis covers only two offences, and the comparison for the second offence is based on a small sample of just 20 sentences. As a result, caution is warranted when generalising these findings to other offences.

Conclusion

As a conclusion, I offer some tentative policy implications. Since sentence inflation is already inconsistent with sentencing law, harmonising suspended and unsuspended prison terms would require no legislative change. Rather, judicial training could place greater emphasis on the distinction between the decision about the sentence duration and the choice of whether to suspend it. A creative application of structured sentencing (e.g. schematic guidelines proposed by Drápal, 2024), could help nudge judges to determine the duration of prison term first, and then decide whether suspending it is warranted. Further good news is that the current inflationary practice can probably be abandoned with no prejudice to the interests that motivated it in the first place (appeasing the public and strengthening individual deterrence). First, the public likely lacks a clear numerical sense of just punishment and might not be sensitive to even substantial differences in the duration of prison terms (Uhl & Pickett,

⁴ Personal communication based on unpublished data obtained from the Ministry of Justice.

2024). Second, the offenders are likewise insensitive to marginal increases in the severity of prison terms (see Pratt et al. 2017 for review)—until the sword of Damocles actually falls upon their heads, they may care little whether it was sharpened beyond its standard upkeep.

References

- Ashworth, A. (1983). *Sentencing and Penal Policy*. Weidenfeld & Nicolson.
- Bagaric, M. (1999). Suspended sentences and preventive sentences: illusory evils and disproportionate punishments. *The University of New South Wales Law Journal*, 22(2), 535-563.
- Bartels, L. (2008). *Sword or Feather? The use and utility of suspended sentences in Tasmania*. University of Tasmania. Thesis.
https://figshare.utas.edu.au/articles/thesis/Sword_or_Feather_The_use_and_utility_of_suspended_sentences_in_Tasmania/23210609?file=40908098
- Bartels, L. (2009). Suspended Sentences-a Judicial Perspective. *Law and Justice Journal*, 9(1), 44-63.
- Bottoms, A. E. (1981). The suspended sentence in England, 1967-1978. *British Journal of Criminology*, 21(1), 1-26.
- Braithwaite, J. (2018). Minimally sufficient deterrence. *Crime and Justice*, 47(1), 69-118.
- Doob, A. N., & Webster, C. M. (2003). Sentence Severity and Crime: Accepting the Null Hypothesis. *Crime and Justice*, 30, 143–195.
- Dölling, D., Entorf, H., Hermann, D., & Rupp, T. (2009). Is Deterrence Effective? Results of a Meta-Analysis of Punishment. *European Journal on Criminal Policy and Research*, 15(1-2), 201–224.
- Drápal, J. (2024). Towards proper sentence justification. In J. Drápal (Ed.), *Justifying Punishment. Perspectives from Post-Communist Europe* (pp. 7–28). Routledge.
- Fischman, J. B. (2013). Measuring Inconsistency, Indeterminacy, and Error in Adjudication. *American Law and Economics Review*, 16(1), 40–85.
- Freiberg, A., & Moore, V. (2009). Disbelieving suspense: Suspended sentences of imprisonment and public confidence in the criminal justice system. *Australian & New Zealand Journal of Criminology*, 42(1), 101-122.
- Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396), 945-960.
- Holste, L., & Spamann, H. (2023). *Experimental Investigations of Judicial Decision-Making: Discussion Paper No. 1096*. Harvard Law School.

- Krajewski, K. (2016). Sentencing in Poland: Failed attempts to reduce punitiveness. *Crime and Justice*, 45(1), 175-219.
- Krajewski, K. (2023). Penal exceptionalism in countries of Central Europe: Why is the region different?. *Archiwum Kryminologii*, 2(XLV), 171-211.
- Morris, N., & Tonry, M. (1991). *Between prison and probation: Intermediate punishments in a rational sentencing system*. Oxford University Press.
- Mycka, K. & Kozłowski, T. (2013). Paradoxy polskiej polityki karnej, czyli jak zapełniamy więzienia nadużywając środków probacji. *Probacja* (2), 5-38.
- Nagin, D. S. (2013). Deterrence in the Twenty-First Century. *Crime and Justice*, 42(1), 199–263.
- Nawój-Śleszyński, A. (2014). Rola środków penalnych związanych z poddaniem sprawcy próbie w kształtowaniu rozmiarów populacji więziennej w Polsce. *Probacja* (2), 5–30.
- Pina-Sánchez, J., & Linacre, R. (2013). Sentence Inconsistency in England and Wales: Evidence from the Crown Court Sentencing Survey. *British Journal of Criminology*, 53, 1118–1138.
- Pina-Sánchez, J., & Linacre, R. (2016). Refining the measurement of consistency in sentencing: A methodological review. *International Journal of Law, Crime and Justice*, 44, 68–87.
- Pratt, T., Cullen, F., Blevins, K., Daigle, L., & Madensen, T. (2008). The empirical status of deterrence theory: A meta-analysis. In F. Cullen, J. Wright & K. Blevins (Eds), *Taking stock: The status of criminological theory* (pp. 367–396). Transaction Publishers.
- Rachlinski, J. J., & Wistrich, A. J. (2017). Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science*, 13(1), 203-229.
- Sommers, R. (2021). Experimental jurisprudence. *Science*, 373(6553), 394-395.
- Spamann, H., & Klöhn, L. (2024). Can Law Students Replace Judges in Experiments of Judicial Decision-Making? *Journal of Law and Empirical Analysis*, 1(1).
- Tait, D. (1995). The invisible sanction: suspended sentences in Victoria 1985–1991. *Australian & New Zealand Journal of Criminology*, 28(2), 143-162.
- Uhl, A., & Pickett, J. T. (2024). The (in) stability of punishment preferences: Implications for empirical desert. *The British Journal of Criminology*.

Wasik, M. (1994). Sentencing guidelines: The problem of conditional sentences. *Criminal Justice Ethics*, 13(1), 50-57.

World Prison Brief. (2024). *World prison brief data*. Institute for Crime & Justice Policy Research.
<https://www.prisonstudies.org/>

van Gelder, J. L., Aarten, P., Lamet, W., & van der Laan, P. (2015). Unknown, unloved? Public opinion on and knowledge of suspended sentences in the Netherlands. *Crime & Delinquency*, 61(5), 669-689.