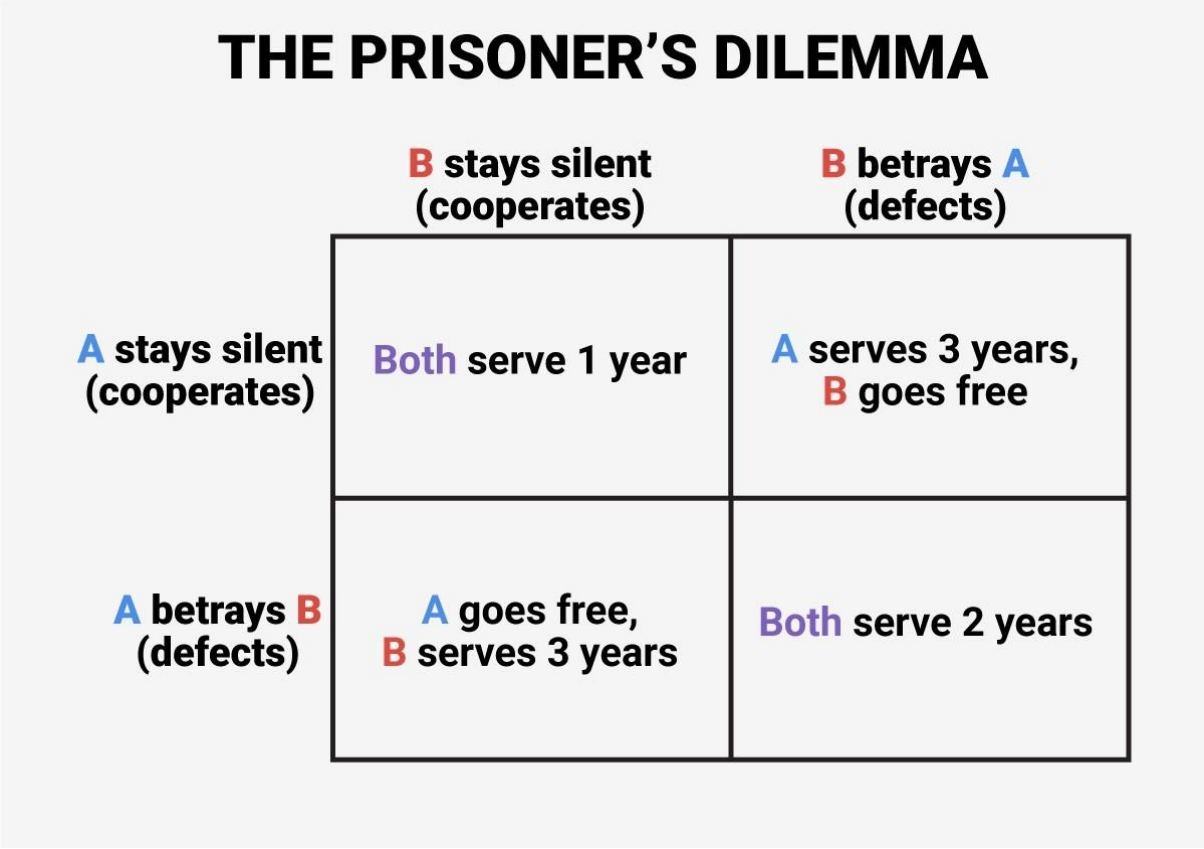
The Prisoner's Dilemma: A Legal Critique

Joseph Wayne Smith, Brittany Smith, and N. Stocks



(History Cafe, 2023)

1. Introduction: The Classical Formulation of the Problem

The prisoner's dilemma in game theory was originally devised by Merrill Flood and Melvin Dresher in 1950, with the familiar formulation of the problem using prison sentences being given by Albert W. Tucker, who coined the name, the "prisoner's dilemma," and expressed the problem in its present form (Dresher, 1981; Tucker, 1983; De Herdt, 2003).

The dilemma can be expressed by considering two suspects, X and Y, who have been arrested, charged, are in custody (no bail to be posted, presumably), facing a trial for some major crime C and a minor crime M. X and Y were allegedly involved in a joint criminal enterprise, committing both crimes C and M. The police prosecutor makes the following deal to each prisoner, who are in separate cells, and cannot communicate with each other, or any third party. Typically, nothing is said about legal representation, so if there are separate lawyers present in police questioning, they are open to the deal, or X and Y have waived the right to a lawyer being present. In any

case the deal recognises that each defendant has the right to remain silent. But, if they choose to confess, and the accomplice remains silent, then all charges will be dropped, and that person walks free, in an immunity agreement (Anonymous, 1981). A variant is that the charge for the major crime C is dropped, but the prisoner still stands trial for M. However, if that offer is accepted then the prisoner is required to testify in the trial as a witness against the accomplice. The police prosecutor believes that this evidence will secure a conviction. But if the accomplice confesses and the other does not, remaining silent, then the same deal applies to them. But if both confess then the case will be made in sentencing for early parole for both prisoners. Since the only evidence in the case is the testimonies of X and Y, if both remain silent, then neither will be convicted of the major crime C, but the minor crime will be pursued, with there presumably being some independent evidence other than the testimonies of X and Y.

The dilemma here is that for both X and Y, whatever either does, confession seems to be a preferrable option to exercising the right to silence. But, if both X and Y confess, then they are both convicted of the major crime C and are worse off than they would be in exercising the right to reman silent. The prisoner's dilemma problem has generated an enormous literature, spanning many fields, as it is a thought experiment illustrating that individual rationality and group/collective rationality can clash (Campbell & Snowden, 1985; Poundstone, 1993; Peterson, 2015). A similar situation is presented in Garrett Hardin's "tragedy of the commons," (Hardin, 1968), where while it may be rational (perhaps in terms of utilitarian maximization), for an individual actor (which could be an individual nation state), to exploit a resource, perhaps to *near* exhaustion, collectively, if everyone acted in the same way, the resource would be totally depleted.

While we do not dispute the claim that here are many unobjectionable formulations of the prisoner's dilemma, as in the case of nuclear weapons deterrence (where the original 1950s research was done), climate change policies, and other areas involving conflicts over resources, we believe that the legal version of the prisoner's dilemma, considered as something even approximating a real-world legal problem, is deeply flawed. The situation is not merely unrealistic, which could be tolerated as a thought experiment; rather, there are essential elements of the formulation of the problem that fly in the face of actual legal practice. Thus, we will now subject the prisoner's dilemma to legal analysis, and show how the dilemma will collapse.

2. Statutory and Common Law Issues

We have already mentioned the oddity of the situation that both X and Y do not seem to have an attorney present during the police questioning, and in reality, the dilemma may not be presented, as it is likely that in a situation where there is no independent evidence of the major crime C, such as forensic evidence, that independent of each

other, the attorneys would be recommending the exercising of the right to silence. They would then need to battle the case the State makes for the minor crime M, but, that's law, and business as usual.

That being said, we now need to look more closely at the police prosecutor's claims. The police prosecutor is claiming that if one of the prisoners confesses and the other does not, that testimony will be used in court as evidence in the trial of the accomplice, as *ex hypothesi*, that is all they have got. The question then is: that is the evidential strength of the testimony of the one who takes the deal, assuming the other does not? Let us say that Y takes the deal, but X remains silent.

We can illustrate the legal aspects of this case using the law of our jurisdiction, Australia, which is more familiar to us than US law, but the issues can be restated there if necessary. Suppose the case takes place in the state of New South Wales. In the situation in which a witness is reasonably supposed to have been criminally involved in events (and the confession of Y is definitely this), section 165 (1) (d) of the *Evidence Act 1995* (New South Wales) provides that evidence that may be unreliable in a criminal proceeding, includes that of "a witness who might reasonably be supposed to have been criminally concerned in the events giving rise to the proceeding." This includes an accomplice as a witness, even without the confession deal, as in the prisoner's dilemma.

The High Court of Australia in the case of *Sio v The Queen* [2016] HCA 32, allowed an appeal against a decision of the New South Wales Court of Criminal Appeal (NSW CCA), on a case involving the unreasonableness of the conviction of the appellant Sio for armed robbery with wounding. The reason for the decision was that the NSW CCA had erred in admitting hearsay evidence (as well as omitting an element of the offence in jury direction).

Sio had driven Filihia to a brothel in Clyde, New South Wales. Filihia, armed with a knife, intended to rob the brothel. However, Filihia in a confrontation with a worker at the brothel, stabbed the worker, killing him. The robbery was completed and Filihia returned to the car that Sio was driving. As an accomplice, Sio was charged with the murder of the worker, and armed robbery with wounding. At Sio's trial, Filihia refused to give evidence, but electronically recorded interviews, and a statement by Filihia were admitted, alleging that Sio was the driver of the car, and had given Filihia the knife.

The evidence was allegedly admitted under an exception to the hearsay rule, section 65 of the *Evidence Act 1995* (New South Wales). Hearsay evidence is an out-of-court statement (s) tendered to the court for the purpose of evidence of the truth of the assertion: section 59, *Evidence Act 1995* (Commonwealth of Australia). Much the same holds in the US Federal Rules of Evidence section 801 (c). If the maker of a

representation is unavailable, and where the representation was against the interests of the person who made it, and that the circumstances made it likely that the representation was reliable, then the evidence can be admitted as an exception to the hearsay rule.

The trial judge directed the jury that they needed to be satisfied that Sio had foreseen the possibility that the knife could be used to commit an offence, as occurred in the murder of the worker. However, the trial judge omitted mentioning this element regarding the armed robbery with wounding charge.

The trial judge held that Sio should be acquitted of murder, but convicted of armed robbery with wounding. The NSW CCA dismissed Sio's appeal that the trial judge did not make an error in admitting Filihia's evidence, and the armed robbery with wounding verdict was unreasonable.

The High court of Australia held that the armed robbery with wounding verdict must be quashed, as a misdirection by the trial judge occurred by erring in admitting the hearsay evidence. The evidence of the representation Sio gave Filihia the knife, was not made in circumstances making the evidence reliable. The High court of Australia rejected the proposal that there should be a new trial in the matter of the charge of armed robbery with wounding, as this would contradict the jury verdict on the charge of murder. As well, the wrongful admission of hearsay evidence precluded substituting a conviction for armed robbery. This left only an order for the charge of armed robbery to be examined in a new trial.

Relevant to our target issue of the evidential basis of the prisoner's dilemma, the High court of Australia said at paragraph [65]:

Evidence by an accomplice against his or her co-offender has long been recognised as less than inherently reliable precisely because of the perceived risk of falsification. Statements by an accomplice afford the classic example of a case where a "plan of falsification" may be expected to be formed, given the obvious interest of one co-offender to shift the blame onto his or her accomplice, especially where the circumstances also include the opportunity to seek to curry favour with the authorities.

This *ratio decidendi* was held in earlier cases in Australia and England: *Peacock v The King* [1911] HCA 66; *Tumahole Bereng v The King* [1949] AC 253, at 265; *Davis v Director of Public* Prosecutions [1954] AC 378, at 391, 399.

Let us now assume that the matter of the *People v. X* goes to trial, based solely upon the Testimony of Y, who has confessed to the crime, but is granted prosecutional immunity. While it is uncertain if specific legislation in the US jurisdiction where the trial occurs mirrors the Australian jurisprudence just discussed, what is clear is that the testimony of Y will be subjected to defense cross-examination. The testimony of Y

must therefore be accepted by the jury as establishing the guilt of X to the standard of beyond reasonable doubt, for a guilty verdict to be delivered. But no unbiased, reasonable jury is likely to conclude that, given the intrinsic unreliability of the testimony, as court authorities we previous quoted have held. A sample of a cross examination now follows:

Attorney for X: Now Mr Y, the jury has heard in the examination your testimony regarding the commission of crimes C and M, based upon your prior confession. And you testified that X was your accomplice in this joint criminal enterprise. To be clear, is that right?

Y: Yes, I said that.

Attorney for X: But isn't it the case that you are now, because of a deal done with the police prosecutor, are not being prosecuted for any crime at all, even though you said you did C and M, as you have testified against X, and now can walk out of this court room a free man. Is that the case?

Y: Yes, I did a deal with them … so what?

Attorney for X: At the danger of asking the obvious, but why did you do the deal if you were as you say guilty? I think the jury would like to know your thought processes.

Y: Well, I was offered my freedom, and why shouldn't I take up the offer? I knew that X would remain silent, so I thought talking was the best thing to do for me.

Attorney for X: So, you were given a golden opportunity and you took it, didn't you?

Y: Sure, why not?

Attorney for X: Are you aware that the State is seeking the death penalty for my client?

Y: Why should I care about that? It is every man for himself nowadays.

Attorney for X: But my client claims to be innocent, and that you did the major crime C, while he was jointly responsible for the minor crime, M. Do you deny this?

Y: X is a bald-faced liar, and as guilty as hell! He deserves the death penalty.

Attorney for X: But why should the jury believe you when you had an incentive to lie, which you are now doing, aren't you?

Attorney for Y: Objection your Honor, the witness is being badgered, and also being asked to speculate.

Trial Judge: Sustained.

Attorney for X: My apologies your Honor; I withdraw the remarks. I have no further questions.

In our opinion the jury would not deliver a guilty verdict solely based upon Y's unreliable testimony. And as we saw in review of *Sio v The Queen*, higher courts are most likely to quash a decision to convict X on the basis of the one piece of evidence seen in the prisoner's dilemma, thus collapsing the dilemma, in its standard legal version, but of course, not in other formulations on these grounds.

3. Objection and Conclusion

It could be objected that these sorts of legal considerations are irrelevant to the soundness of the prisoner's dilemma, which after all, is a logico-conceptual problem in game theory. That is true as far as it goes. Certainly, the prisoner's dilemma has many versions that do not have a legal framework at all, as we have emphasized throughout this essay. That we grant. But, the point made here is that the presentation of problems should, even if idealized, have a reasonable degree of theoretical and empirical/practical accuracy to real world situations. In the case of the standard legal interpretation of the prisoner's dilemma, the actual real world working of the law, at least in the common law jurisdictions, undercuts the problem, because the legal premises used to set up the problem, while interesting as a thought experiment, are not only merely unrealistic, but also have a self-undermining aspect to them, since the very evidence that is supposed to create the problem, undermines the problem.

REFERENCES

(Anonymous, 1981). Anonymous. "Accomplice Testimony and Credibility—'Vouching' and Prosecutional Abuse of Arguments to Testify Truthfully." *Minnesota Law Review* 65 (6):1169-1186.

(Campbell & Snowdon, 1985). Campbell, R. & Snowden, L. *Paradoxes of Rationality and Cooperation*. Vancouver BC: Univ. of British Columbia Press.

(De Herdt, 2003). De Herdt, T. "Cooperation and Fairness: The Flood-Dresher Experiment Revisited." *Review of Social Economy*, 61 (2): 183-210.

(Dresher, 1981). Dresher, M. *The Mathematics of Games of Strategy: Theory and Applications*. Englewood Cliffs NJ: Prentice-Hall.

(Hardin, 1968). Hardin, G. "The Tragedy of the Commons." *Science* 162: 1243-1248.

(History Cafe, 2023). Marathe, Y. "The Prisoner's Dilemma." *History Cafe*. Available online at URL = <https://historycafe.in/the-prisoners-dilemma/>.

(Peterson, 2015). Peterson, M. (ed.). *The Prisoner's Dilemma*. Cambridge: Cambridge Univ. Press.

(Poundstone, 1993). Poundstone, W. *Prisoner's Dilemma*, New York: Anchor.

(Tucker, A. W. (1983). "The Mathematics of Tucker: A Sampler." *The Two-Year College Mathematics Journal* 14 (3): 228-232.