

Open in app ↗



Search

Write

Photo by [Bryson Hammer](#) on [Unsplash](#)

✦ Member-only story

# Pushing Back: Lessons from the Prisoner's Dilemma

Apply the IPD Solution to Employee/Manager Relationships



Kenneth Kousen

Published in The Pragmatic Programmers · 7 min read · Jul 23, 2021



187



2



The Prisoner's Dilemma from mathematical game theory makes for an interesting launching point for a discussion about cooperation. The *Iterated Prisoner's Dilemma* (IPD) is even more interesting, and you can apply a solution for IPD to the employee/manager relationship. In fact, IPD can help you think about interactions with your manager as a non-zero-sum game — meaning both sides can win.

In the classic Prisoner's Dilemma, two prisoners, A and B, are captured for their roles in the same crime. The police interrogate each separately, and neither prisoner can communicate with the other. Each has a choice of whether to cooperate with the other by staying silent, or defecting and turning in their partner. What can happen looks something like this:

	Prisoner B cooperates (by staying silent)	Prisoner B defects (by betraying)
Prisoner A cooperates (by staying silent)	Each serves time (1 year)	Prisoner A serves time (3 years) Prisoner B goes free
Prisoner A defects (by betraying)	Prisoner A goes free Prisoner B serves time (3 years)	Each serves time (2 years)

If you **play the game only once**, it's easy to determine what to do. Say you are prisoner A. You look at B's options and see which move is best for you.

- If B decides to *cooperate*, then if you cooperate, you serve a year, but if you defect you go free, so *defect is better*.
- Say B decides instead to *defect*. Then if you cooperate, you serve three years, but if you defect you serve only two years. Again, *defect is better*.

In either case, defecting minimizes your losses no matter what B does, so the best strategy is to defect. Of course, the situation is symmetrical, so if B goes through the same analysis, B will defect as well. It's a shame that the most rational outcome is that both parties will turn on each other, but that's what the math says.

*Incidentally, this situation is a classic example of a Nash equilibrium, in case you enjoyed the movie A Beautiful Mind.*

## The Iterated Prisoner's Dilemma

The situation changes when you play the game over and over again an indefinite number of times. That's the Iterated Prisoner's Dilemma, and the successful strategies are no longer so obvious.

*If you're interested, Wikipedia has a nice summary of the problem and common strategies for it here.*

Back in the 1980s, a researcher named Robert Axelrod held a series of computer tournaments to investigate strategies to win the IPD problem. He summarized his results in a book called The Evolution of Cooperation, which is considered a classic in the field. The tournaments he held were dominated by a program called *Tit for Tat* (TFT), a four-line program written in BASIC (!) that consisted of the following strategy:

1. Cooperate on the first move.
2. Every move after, echo what the opponent did on the previous move.

In other words, the TFT program would start by cooperating, and then if the opponent cooperated, it would do so again, but if the opponent defected, it would do that on the next move. This strategy succeeds because it:

- Starts off by cooperating.
- Never defects first.
- Retaliates immediately in the case of a defection.
- Forgives immediately if the opponent resumes cooperating.

We call this a *nice* strategy. All high-scoring strategies start with cooperation.

If you would like to experiment with the IPD game and see various strategies in action, try [this excellent online simulator](#). Note that the simulator refers to TFT as *Copypcat*, but it's the same strategy. As a (mild) warning, the website plays music in the background, so you might want to turn that off if you're back in the office 😊.

## Applying TFT to Your Manager



Photo by [Jason Leung](#) on [Unsplash](#)

The IPD problem has been applied to many situations, from environmental studies to economics to international politics. Here we'll apply it to the employee/manager relationship.

One important idea to remember is that you're not trying to *defeat* your boss or win the game in any absolute sense.

You need your manager as an ally within the organization, so if you're in an actual serious conflict with your manager, even if you win, you lose. That's why non-zero-sum games are so important. You need to find a way for both sides to get what they want if you're going to build a sustainable relationship.

The key to the TFT strategy is that cooperation emerges naturally as long as both sides remember you're going to do this again, and again, and again.

Whenever one side defects, it affects how the other side reacts next time. It's therefore in the best interests of both sides to find a way to work together.

The steps to apply TFT to your manager are:

1. Start by *going along* with your manager's decisions.
2. When they do something you don't like, *push back* in a way that does not threaten the constructive loyalty relationship. In other words, shut the door. No one other than the two of you needs to know that you're having this disagreement. Talk to them privately and say why you're not happy, without becoming too emotional. The goal is to be heard, and for them to understand this is important, but not a crisis.
3. *Negotiate* a resolution, which may or may not get you anything.
4. In any case, *go back to work*.

Your objections need to be timely because you want your manager to connect what they did with why you were unhappy. You can't wait six months and then blow up, or you risk your reaction being discounted as being way out of proportion to whatever is going on at the moment. A brief, calm objection right away is much more valuable than a huge argument later.

The last step, going back to work, can be hard to accept. You're trying to communicate that the current problem is an issue, but not a crisis. Your boss needs to know that you will object when you're not happy, but you're also making it clear that this is not the end. That may feel like you're undercutting your own objections. After all, if you're just going to go back to work anyway, does that mean your boss got away with whatever they did?

Maybe, but even so, it's still worth it, for a couple of reasons:

- Simply making the objection will have an impact on most managers. Most managers are people too (despite evidence to the contrary), and they don't enjoy conflict any more than the rest of us.
- Objecting also means standing up for yourself. You're making it clear that what you feel is important and worth hearing, even if nothing is done about it this time. That will feel much better than simply swallowing your feelings as being not worth expressing. You're also learning that you can push back without turning the situation into a crisis, and that's validating.

Again, you're not trying to win this particular battle. You're training your manager. You want them to understand that if they do something again that goes against your interests, they're going to hear about it again, and again the time after that, and again after that, too.

## Managers Who Refuse To Learn



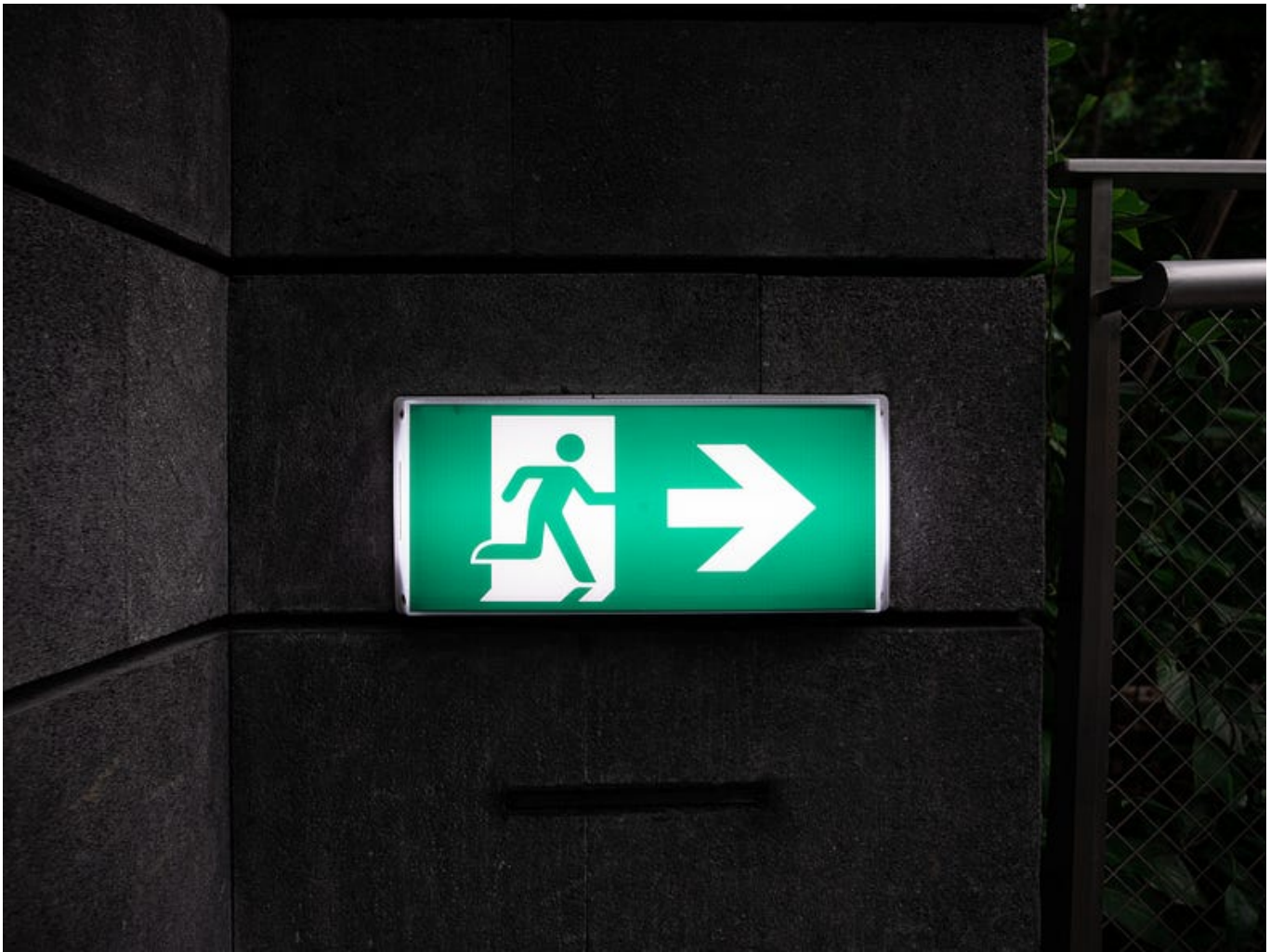


Photo by [Andrew Teoh](#) on [Unsplash](#)

What if, after making your objections, your manager doesn't get any better? They don't apologize or even acknowledge your objections were valid, and they continue the behavior you don't like.

What then?

Remember, you're playing the long game. Your hope is that the boss will get tired of the objections.

They'll also get to know better what you care about enough to come to them and push back. Eventually, hopefully, they'll learn to keep in mind what you want when they make decisions that affect your career.



Some managers never learn, however. Either they don't agree they're wrong, or they aren't capable of changing their behavior, or they just don't care enough to try to accommodate you. After all that, then it really is time to leave.

The good news, however, is that if it comes to that, neither side will be surprised. You never want your announcement that you're leaving to be a surprise to your manager. If that happens, both sides missed their chance to make the situation better.

If you do eventually decide to leave, the first question HR will ask you at the exit interview is, "Did you tell your manager you weren't happy?" You really want to be able to say *yes* to that.

Besides, some managers really do learn.

## More Wisdom From *Help Your Boss Help You*

If you enjoyed this article, you may also enjoy the following articles, which are based on the concepts from *Help Your Boss Help You* by Ken Kousen:

Hacks: Your Boss Is Not Your Friend

What Do Popular TV Shows Have to Do with Your Work Relationships?

medium.com

Answering the Unanswerable