Doug Moore Aug 14, 2017 4 min read

The Evolution of Trust: this adorable game explains the math behind interpersonal trust

Cooperation and trust are two of the most important elements of any successful human endeavor — people need to work together to accomplish big things, and mutual trust helps to avoid squabbling and backstabbing. Despite their self-evident value, of course, these social bonds between individuals and groups break down all the time in practice. This basic fact of human relationships raises an important question: how can we create social conditions that encourage trust and cooperation?

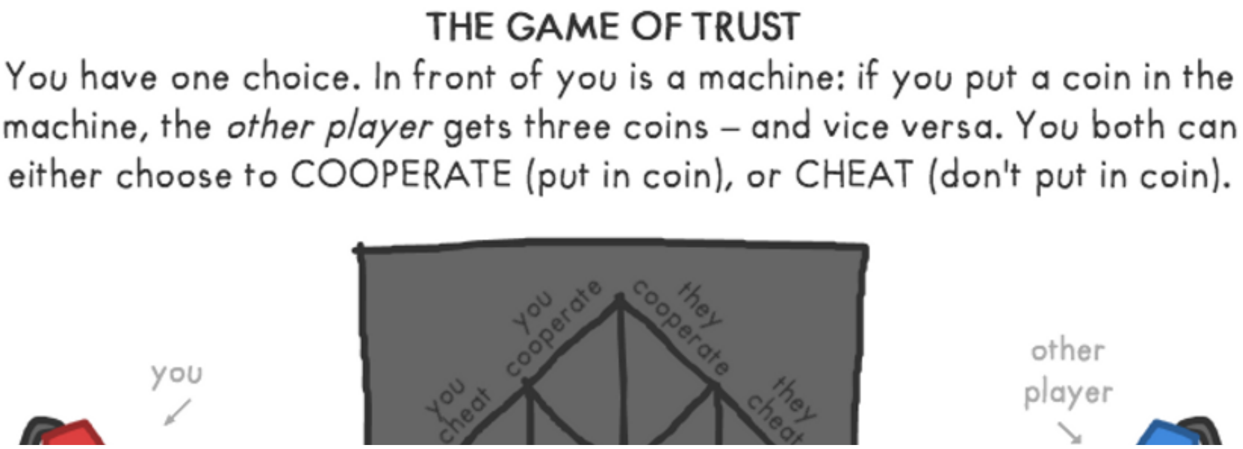
Game theory has been considering this problem for decades, and has produced some valuable and surprisingly clear-cut insights into the matter. "The Evolution Of Trust," a brief and free-to-play online game developed by the educator and game designer Nicky Case, provides an excellent overview of these insights and the logic that underlies them. Based primarily on Robert Axelrod's seminal 1984 book The Evolution Of Cooperation, "The Evolution Of Trust" distills some knotty concepts from game theory down into a cute, entertaining half-hour game experience.

We strongly recommend playing the game (and, if you like it, considering a donation to Case's Patreon). But if you don't have the ability or inclination, we've summarized a few of its most interesting takeaways below. Here's the bite-sized version; keep reading for the full summary:

Establishing trust and cooperation between potential competitors involves meeting 3 requirements:

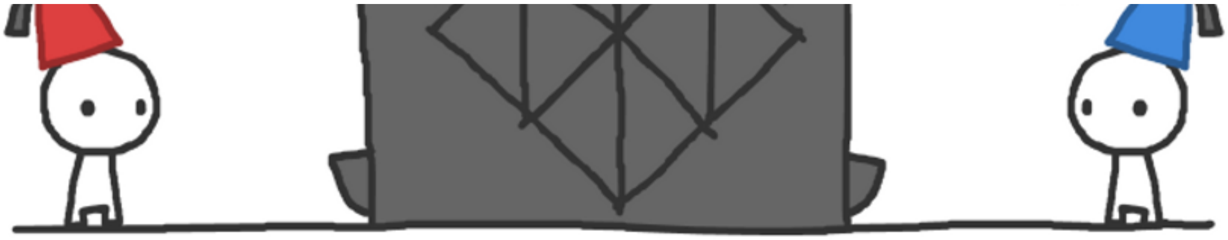
1. *Repeat interactions*
2. *Possible win-wins*
3. *Low miscommunication*

In order to model situations where trust is important, "The Evolution Of Trust" uses a very simple two-player game. The game looks like this:



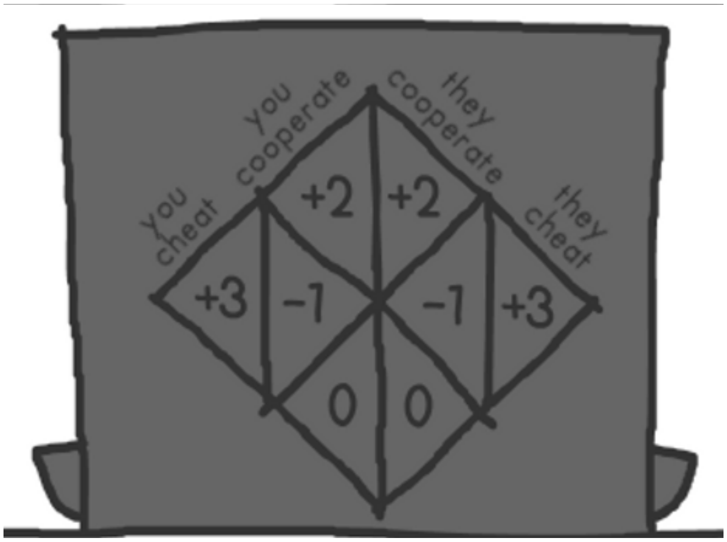
THE GAME OF TRUST

You have one choice. In front of you is a machine: if you put a coin in the machine, the *other player* gets three coins – and vice versa. You both can either choose to COOPERATE (put in coin), or CHEAT (don't put in coin).



If you've studied game theory before, you probably recognize this game as a variation on the famous Prisoner's Dilemma game. In such games, both players stand to benefit if they both choose the "cooperate" option, but choosing the "cheat" option when the other player chooses "cooperate" is even more beneficial — and your opponent takes a hit. However, if both players "cheat," neither benefits at all. Consequently, in order for both players to do well, they must resist the temptation to cheat, trust each other, and cooperate.

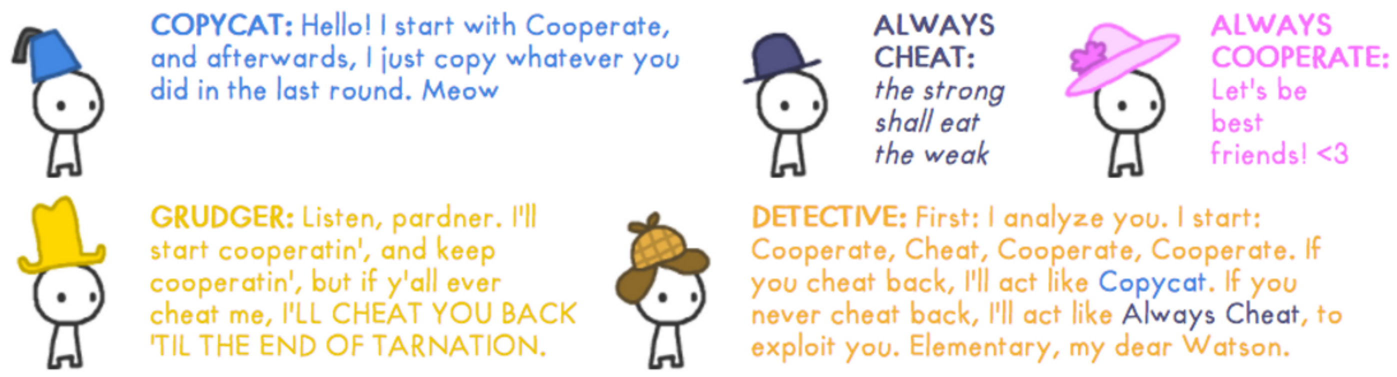
The specific payoff scheme in "The Evolution Of Trust" looks like this:



"The Evolution Of Trust" explores the implications of this game in 3 separate scenarios:

- Individual 1-on-1 games, with 10 rounds per match
- Round-robin tournaments in which multiple different players each play against each other once in 10-round matches
- Repeated (or "iterated") tournaments, in which numerous players using a variety of strategies face off against each other multiple times; in these repeated tournaments, the 5 lowest-ranking players are eliminated and the 5 best-performing players are duplicated after every round, so that players using the most effective strategies slowly take over the game

The players that participate in these tournaments are also representations of famous Prisoner's Dilemma strategies from the game theory annals, including the following:



COPYCAT: Hello! I start with Cooperate, and afterwards, I just copy whatever you did in the last round. Meow

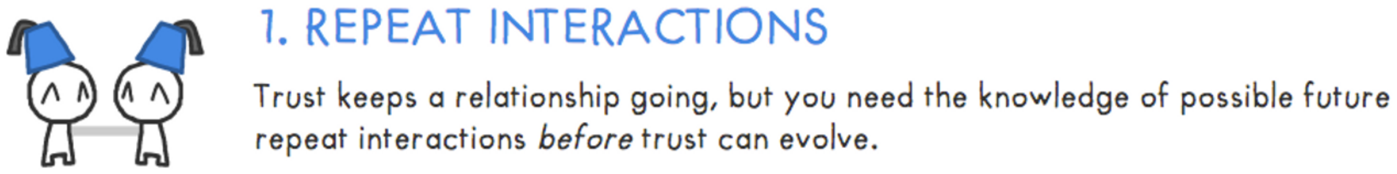
ALWAYS CHEAT: *the strong shall eat the weak*

ALWAYS COOPERATE: Let's be best friends! <3

GRUDGER: Listen, pardner. I'll start cooperatin', and keep cooperatin', but if y'all ever cheat me, I'LL CHEAT YOU BACK 'TIL THE END OF TARNATION.

DETECTIVE: First: I analyze you. I start: Cooperate, Cheat, Cooperate, Cooperate. If you cheat back, I'll act like Copycat. If you never cheat back, I'll act like Always Cheat, to exploit you. Elementary, my dear Watson.

The game plays out a number of variations on the basic scenarios outlined above, using each version of the game and participating player strategies to illustrate a different essential condition for the development of trust among rival players. It names three such conditions in all:

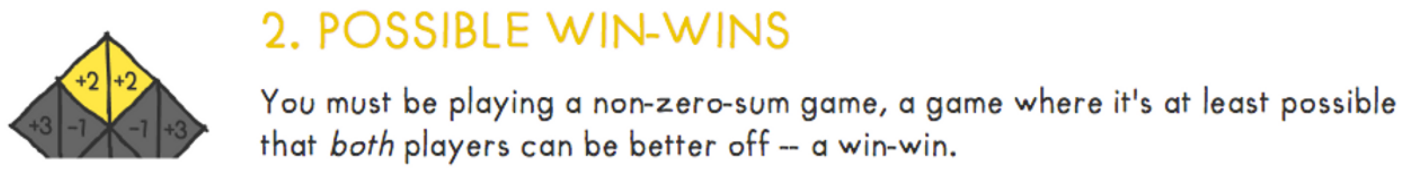


1. REPEAT INTERACTIONS

Trust keeps a relationship going, but you need the knowledge of possible future repeat interactions *before* trust can evolve.

In individual games, the best move is typically to choose the "cheat" option in every round — you don't know what kind of person your opponent is and you'll never have to deal with them again, so your best bet is to try to screw them over before they can do the same to you. (In Prisoner's Dilemma games, this is a mathematical fact!) But if your fate is determined by multiple repeated games, burning every bridge you come across can come back to bite you, and learning to work with others can provide long-term dividends.

"The Evolution Of Trust" illustrates this with its most basic repeated-tournament illustration. Players using the Always Cheat strategy tend to win single-round Prisoner's Dilemmas, but over multiple rounds, the Copycat strategy — which plays nicely with benign strategies like Always Cooperate, but viciously retaliates against hostile strategies like Always Cheat — tends to prevail. As the game points out, this provides a certain mathematical substantiation for the famous Golden Rule of ethical conduct.

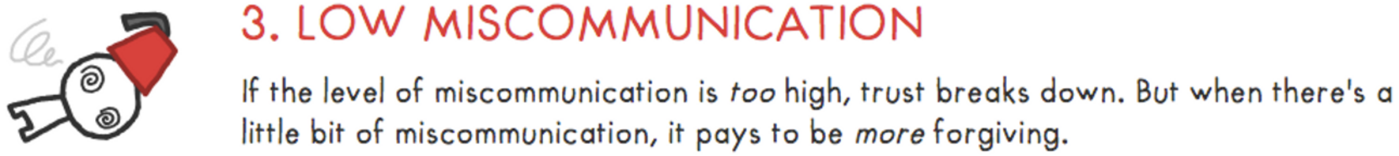


2. POSSIBLE WIN-WINS

You must be playing a non-zero-sum game, a game where it's at least possible that *both* players can be better off -- a win-win.

In the standard Prisoner's Dilemma game depicted by "The Evolution Of Trust," cooperation between players tends to improve both of their lots. That means it's a non-zero-sum game — a game in which both players have the possibility of coming out ahead if they play their cards right.

But not all games work this way — not even all versions of the Prisoner's Dilemma game. If the payoffs for cooperation are reduced, or the payoffs for cheating are increased, the game becomes zero-sum — the possibility of mutual benefit disappears, and the incentives towards trust disappear with them. "The Evolution Of Trust" features a version of the iterative tournament where the Prisoner's Dilemma payoffs have been tweaked in this fashion, causing the brutal Always Cheat strategy to win out over multiple tournaments.



3. LOW MISCOMMUNICATION

If the level of miscommunication is *too* high, trust breaks down. But when there's a little bit of miscommunication, it pays to be *more* forgiving.

In the real-world situations that the Prisoner's Dilemma is designed to model, accidents happen sometimes — and those accidents can send the wrong signal to the other player. ("The Evolution Of Trust" illustrates this idea by showing a player approaching the machine to drop a coin in, but tripping and falling down at the moment of truth, which the other player perceives as a deliberate "cheat" choice.)

"The Evolution Of Trust" illustrates the implications of this problem nicely. Strategies like Copycat can backfire if someone makes a mistake, setting off an endless cycle of vengeance. When mistakes are relatively rare, it's actually beneficial for rule-based strategies like Copycat to become more forgiving and less likely to retaliate. (The game features a modified version of the Copycat designed to illustrate this idea, called the Copykitten.) But if mistakes are common — with a 10% likelihood or greater, in this particular version of the Prisoner's Dilemma — hostile trust-free strategies like Always Cheat tend to prevail in the long run.

"The Evolution Of Trust" concludes with an important point that transcends the Prisoner's Dilemma: the way people treat each other in game-like situations depends heavily on the nature of the game itself. We can give people a better chance of getting along in these situations by setting up incentives that make trust and cooperation attractive. (Though, as Case's footnotes remark, this tends to be more easily said than done.)

Once again, we recommend that you play through the game itself — it's short, simple, and gives you a lot of opportunities to play around with these concepts yourself.