

It seems we would all be better off if we cooperated and did what's best for the group, but that often requires us to bear costs individually. Explaining this challenge has been a central focus of the natural and social sciences which have studied reciprocity with simple behavioral models. However, these models cannot account for a key feature of human social life: the scale, scope, and flexibility of our cooperation. Here, we give a reverse-engineering account of human cooperation grounded in inference and abstraction rather than behavioral reflex. Our model robustly cooperates and learns who to cooperate with from just sparse, noisy, and overdetermined observations of behavior. Like people, it does this even when it never plays the same game nor interacts with the same person twice. Finally, our model decisively outcompetes existing behavioral approaches even in the settings the behavioral strategies were originally optimized for.