The neural mechanisms of learning to balance fairness and self-interest: a reinforcement learning account

While norms often act as useful heuristics, norms sometimes prevent the exploration of valuable courses of action. In the current study we combined computational modeling, the ultimatum game, and fMRI to explore the behavioral and neural correlates that accompany circumstances in which social norms impede optimal decision-making. The ultimatum game is a dyadic paradigm in which one player, the proposer, decides how much of an endowment to offer a responder, who decides whether to accept the offer, in which case the endowment is divided as proposed, or reject the offer, in which case both parties receive nothing for that trial. We examined the behavior and neural activity of proposers playing against groups of responders with different acceptance functions as well as computer generated lotteries programmed to mimic human behavior. Our goals were (i) to assess whether or not proposer behavior and neural activity could be captured with a reinforcement learning framework, and (ii) to see if this learning process differed between social and non-social conditions. Using a novel reinforcement learning algorithm, we found proposer behavior to be governed by subjective representations of the opponents’ acceptance thresholds. Specifically, we found that proposers learned the slope and intercept of each opponent’s acceptance logistic function, which mapped the investment amount into a victory or a defeat. We furthermore found that learning differed between the social and non-social conditions, with subjects learning faster in the non-social condition, and exhibiting an apparent reluctance to explore the acceptance range of the opponents in the social condition.