Learning fairness and the rules of engagement in social interactions

Global challenges require individuals from different groups to interact and agree on a joint course of action. This can be difficult because what is deemed fair and appropriate in one group may be considered unfair and offensive in another. Here we model social interactions as ultimatum bargaining and elucidate how humans as proposers learn a priori unknown rules of engagement in responders from other groups. We develop a computational model of individuals as Bayesian learners with inequity aversion, which predicts that individuals behave more generously than is sometimes needed, in turn forming biased beliefs about ‘what it takes’ for responders to agree. Behavioral and neuro-imaging experiments confirmed those predictions and mapped biased learning variables in brain regions associated with reinforcement learning and social decision-making. Initial generosity can be personally costly and create incorrect beliefs about others’ rules of engagement, yet also facilitates cooperation and agreement across group boundaries and cultural divides.