from trust in groups to trust in individuals

@philippparnamets
Philip Pärnamets, Tobias Granwald & Andreas Olsson
philip.parnamets@nyu.edu

background & aim
- trust is central for social interaction.
- recent work investigated how we learn to trust other individuals by tracking their average trustworthiness using basic learning mechanisms.
- however, our social world is structured by group information and little is known about how such information affects, if at all, decisions to trust.

we investigated how trust in individuals is learned and unlearned in the context of minimal social groups.

we aimed to characterize the dynamics both of trial-by-trial updating of trust (learning) and of within-trial variation in trust decisions.

procedure

- trust games against computerized opponents
- groups identified by t-shirt color
- one group reciprocated with P=.75 and other group with P=.25.
- task incentivized
- two phases of experiment (see below)

learning phase
- (learn about groups)
- multiple trust games
- trustworthiness rating of groups
- selection of group for another round

character phase
- (learn about individuals)
- repeated trust games
- trustworthiness rating of individual partners
- selection of group for another round

RL model
- hybrid PH-RW model
- decision: $\delta'$(trust) = $P_h = \alpha \delta'_{0} + \beta (\delta'_{0} - \delta'_{1})$
- learning: $\delta'_{0} = \delta'_{1} + \delta'_{2}$
- $P_{1} = P_{h} \times \exp(\delta'_{2})$
- $P_{0} = 1 - P_{1}$

results – learning to trust

- participants rapidly adjust to actual behavior of partner
- no difference in based on group membership on final trial
- at start of character phase participants discount prior learning (captured by $\kappa$)
- higher learning rate in character phase to negative prediction errors compared to positive ones

results – deciding to trust

- posterior predictive checks
- participants’ choices & RTs can be explained by drift-diffusion model using quantities derived from RL model
- trustworthiness estimates affect both drift rates and starting points of sampling process
- looking behavior during deliberation has additive effect on drift rates
- certainty in partners’ trustworthiness and pupil-linked arousal regulates boundaries of decision process

conclusions

- group information (stereotypes) can be quickly overcome given individuating information in available, with minor long-term effects on aggregate behavior.
- participants compute expectations of reciprocation (as predicted by RL model) and use these to guide both behavior and judgments.
- combination of discounting of prior learning together with faster learning rates for negative prediction errors allow participants to avoid myopia and effectively manage learning.
- participants’ trust decisions depend on trial-by-trial variations in visual attention and pupil-linked arousal.