

Trust-ya: design of a multiplayer game for the study of small group processes

Jerry Huang, Joshua Jung, Neil Budnarain, Benn McGregor and Jesse Hoey

David R. Cheriton School of Computer Science

University of Waterloo

Waterloo, Ontario, Canada

Email: {j354huan,j35jung,bbudnara,bamcgreg,jhoey}@uwaterloo.ca

Abstract—This paper presents the design of a cooperative multi-player betting game, *Trust-ya*, as a model of some elements of status processes in human groups. The game is designed to elicit status-driven leader-follower behaviours as a means to observe and influence social hierarchy. It involves a “Bach/Stravinsky” game of deference in a group, in which people on each turn can either invest with another player or hope someone invests with them. Players who receive investment capital are able to gamble for payoffs from a central pool which then can be shared back with those who invested (but a portion of it may be kept, including all of it). The bigger gambles (people with more investors) get bigger payoffs. Thus, there is a natural tendency for players to coalesce as investors around a “leader” who gambles, but who also shares sufficiently from their winnings to keep the investors “hanging on.” The “leader” will want to keep as much as possible for themselves, however. The game is played anonymously, but a set of “status symbols” can be purchased which have no value in the game itself, but can serve as a “cheap talk” communication device with other players. This paper introduces the game, relates it to status theory in social psychology, and shows some simple simulated and human experiments that demonstrate how the game can be used to study status processes and dynamics in human groups.

I. INTRODUCTION

In games research, it is often desirable to design a mechanism such that agents acting selfishly will produce results with beneficial properties, like the maximization of overall welfare, or the equality with which wealth is distributed. However, in the real world, we may be forced to participate in ‘games’ that tend toward undesirable outcomes, either due to their design, or the irrationality of their human players. It is therefore important to study games that allow players to act against the common good, and to understand how the addition of new agents can shift their behaviour.

Such games can often be cast as social dilemmas, such as the ‘Bach or Stravinsky’ (BoS) problem.¹ In its basic form, BoS is a 2×2 matrix game that may be summarized as follows:

Two friends are interested in going to the symphony. One prefers Bach, while the other prefers Stravinsky, but it is most important to both that they go together.

The dilemma is clear, since both agents cannot have their preferred option while still being together, and one must settle

¹In other literature, this problem may sometimes be referred to as the ‘battle of the sexes’ or the ‘shopping/football problem’ [1].

for the other’s preferred option in order to maximize utility. In this paper, we present a new game, called *Trust-ya*, that generalizes BoS to N players and incorporates elements that expand the space of strategies, such as limited communication between players through the use of arbitrary symbols (emoji). The objective in the design of this game is to provide a platform for small-group research that captures the notions of status and durable inequality [2], [3], [4], while remaining constrained enough to be approachable from a modeling perspective. According to Ridgeway [4], status is the esteem and influence granted by others, distinct from direct power or wealth. It is a kind of inequality that imposes a hierarchy on individuals, as well as groups within a society. In *Trust-ya*, emojis serve as status symbols, and the patterns by which they are accumulated allow us to observe how a social hierarchy can be formed and manipulated.

We describe the game, introduce simulated agents that map out the space of baseline gameplay, and give illustrative results from our (human) design team playing the game. Even in these simple demonstrations, we see the emergence of interesting and expected status behaviours. We conclude with a discussion of the social psychological background to the game, and present a mapping between game and theoretical elements.

Ultimately, it may be possible to analyze group behaviours *a posteriori* in order to evaluate why a group failed to find a collaborative equilibrium, or to suggest mechanisms to achieve an equilibrium from an existing non-collaborative state. For example, such analysis may suggest that a particular kind of artificial agent be inserted in a way that induces an increase in efficiency. However, further research and analysis is needed on the dynamics, effects and ethics of such interventions.

II. METHODS

Trust-ya is a game in which 3 or more players aim to maximize their personal stash of coins drawn from a large central pile. Although they are in competition for these coins, players must cooperate to draw as many as possible from the central pile before the game ends. At each game step, a player may choose to either take C_{take} coins from the central pile for themselves, or give C_{give} coins from the central pile to any other player, where C_{take} and C_{give} are quantities that may be varied by the experimenter. Any player who receives coins (optionally, above some minimum threshold) may “invest”

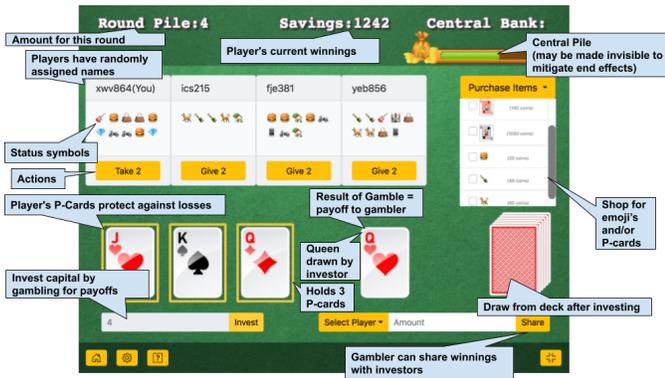


Fig. 1. Annotated game interface at a central point in a game.

those coins and obtain a payout as a function of the invested amount. This function is scaled such that one player investing x coins can expect a greater average payout than n players each investing x/n coins. The investing player may then decide how to distribute this payout among those players who contributed. This process is repeated until a stopping condition is met (most commonly, when the central pile is empty).

Between rounds, players may spend some of their accumulated coins to purchase “P-cards” (which provide investment benefits, as described in Section II-A), or “status symbols,” which have no intrinsic value, but serve as the only medium of communication between players. These symbols are currently displayed as emoji next to a player’s name (from the set (🍔 🍷 🐶 🎮 ... 🛍 📱 🏠 🏡)) but could be represented in a number of ways. The game interface is shown in Figure 1 as a snapshot at about the mid-way point of a game.

A. Game details

For a group of N players, the central pile initially contains $C_{ppp} \cdot N$ coins, where C_{ppp} is the per-player pot size. Each player begins with an empty savings pile, and can see the names and status symbols of all players. Other information, like each player’s accumulated wealth, remains hidden. After all players have chosen to give or take coins in a round, they are each able to see who gave coins to them, but will only see the investment results of any player to whom they gave coins.

When a player makes an investment, a card is drawn from a standard 52-card deck. For non-face cards, the amount received increases with the size of the investment and the value of the card drawn, according to the function:

$$\text{pay}(i) = \text{round}[\max(2^{i/2} - 1, 0) \cdot v \cdot \alpha] \quad (1)$$

where i is the number of coins invested, v is the value of the card and $\alpha = 50/N$ is a scaling factor which is decreasing with the number of players in the game. The scaling factor is used in order to allow the game to proceed much faster in the case of very small games, as in these cases the network dynamics tends to stabilize quickly, and the progression of the game becomes unnecessarily slow without it. This function is such that winnings increase exponentially with the investment, which incentivizes players to attempt to form as large a group as possible. Further, the stochastic nature of drawing cards

mimics uncertainty in real-life investments, and is necessary to introduce variance in payouts. Without such variance, it is too easy for players to make unthinking, repetitive decisions.

If a face card is drawn, a payout is only received if the investor possesses the corresponding “P-card” (protection card), which can be purchased between rounds with saved coins. There are three such cards, one for each face card, where the cost of the Jack P-card is lower than the Queen, which is lower than the King. If the investor has a P-card corresponding to the drawn card, then they receive a payout according to Equation 1, where v may be set by the experimenter, but should generally follow the convention $v_{jack} \leq v_{queen} \leq v_{king}$.

If the player draws a face card while not in possession of the corresponding P-card, they get no payout and additionally lose a number of saved coins determined by the function:

$$\text{loss}(r, i) = \text{round} \left[(s + r - i) \cdot \frac{i}{r} \cdot \frac{v}{100} \right] \quad (2)$$

where r is the number of coins that were available for the player to invest that round (distinct from i , the number actually invested), and s is the size of the player’s savings pile. All coins lost are returned to the central pile. By this formula, loss is determined by the fraction of coins bet in a round, rather than the absolute number, which avoids any incentive to attract fewer investors to reduce risk.

Once payouts have been determined for all investors, each must decide how to distribute the coins gained among their backers (i.e. all players who gave them coins that round). Buttons are provided for quickly sharing the wealth evenly, or sharing nothing, and any other split may be entered manually. At this stage, P-cards may be lost by players with few backers. P-cards are lost due to a lack of influence, based directly on the number of supporters a player has.

The game ends either when the central pile has been depleted, or sometime after a predetermined number of rounds has passed. In order to minimize the impact of a fixed number of rounds on strategy, there is a chance of termination on every turn after the limit. For the remainder of this paper, we will assume a round limit of 50, with a termination chance of 10%.

B. Measuring Utility and Equality

We evaluate the final outcome of a game by the total utility achieved among all players, and the equality with which that utility is distributed. Total utility is given by the total number of coins gained from the central pile, while equality is measured by the Gini coefficient.² The Gini coefficient takes a minimum value of 0 when wealth is distributed completely evenly, and a maximum value of $1 - 1/N$ when one player hoards all wealth, and the remaining $N - 1$ players receive nothing. For a Trust-ya, we define a “good” outcome to be one where total utility is maximized, and the Gini is minimized.

III. RESULTS

A. Simulations

We first ran a set of agent-based tests on the game to establish a baseline for different levels of earnings and equal-

²Gini is as a metric for measuring wealth or income inequality [5].

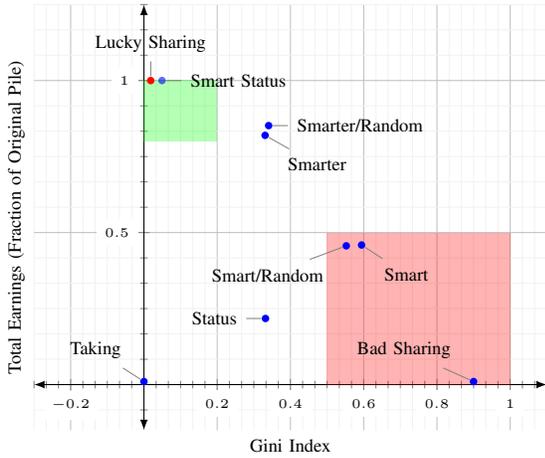


Fig. 2. Equality (Gini) and Wealth (Total Earnings) for simulations with a single agent type (see text). Green area indicates best outcomes on both dimensions (to within 20% of the maximum possible earnings, and $G \leq 0.2$). Red area indicates worst outcomes (50% of the maximum, and $G \geq 0.5$). The optimal bot \bullet for homogeneous populations is shown.

ity, as shown in Figure 2. These simulations used values $C_{ppp} = 10000$, $C_{give} = 2$, $C_{take} = 2$, $v_{jack} = 10$, $v_{queen} = 25$, $v_{king} = 50$.

While (*Bad Sharing*, *Lucky Sharing*, and *Taking*) used an artificial coordination mechanism, the others did not. All agents take every round in *Bad Sharing*, leading to a very equitable, very inefficient payoff. *Taking*, in which all give to one randomly chosen, who never invests, shares or purchases, does badly on both counts. *Lucky Sharing* is the same but the lucky player invests everything, shares evenly, and buys P-cards. This is the best possible solution in terms of efficiency and equality. The non-coordinated bots with settings *Smart*, *Smart/Random*, *Smarter*, and *Smarter/Random* use a simple reinforcement learning strategy, giving coins to the bot with the best perceived rate of return based only on interactions with them. *Smart* bots buy P-cards when they can, *Smarter* bots only buy P-cards if they estimate they will be able to keep them, while the versions with *Random* in them give their coins to a random other player 5% of the time. While the randomness makes little difference, likely due to the limited length of the games, the reticence for buying P-cards helps with earnings, and reduces inequality. *Status* and *Smart Status* bots attempt to invest in the player with the most emojis/status symbols, and are also not coordinated. A *Status* bot will attempt to buy an emoji every round, while a *Smart Status* bot will only do so if it received coins from at least one other player in the last round. *Status* bots perform poorly, as a leader cannot be distinguished when all players have the same number of emoji, but the *Smart* version do not have this problem, and so achieve high total utility and low inequality.

These bots simulations are done with all bots playing the same strategy to establish these as baselines. More complex interactions may be simulated, leading to a myriad of combinations in which different strategies interact to produce different outcomes. It is precisely these more complex interactions that form the basis of the emergence of leaders in the human tests (see next Section).

B. Human Pilot Tests

A team of researchers from the Computational Health Informatics Laboratory at the University of Waterloo were part of the overall research project, including the authors and 5 others including one postdoctoral fellow, three undergraduate students, and two graduate students, assisted in a number of design phases on a rolling basis. The number of participants in our design sessions varied from 4-14. During each session, we briefed all participants on any changes to the game since the last time. We then played a single game until the outcome of the game was clear or time ran out (approx. 1/2 hour). All participants were debriefed, and then we drew conclusions about next design steps. We engaged in about 6 of these sessions, and then ran two fully anonymized sessions of 1/2 hour each with 7 and 8 participants respectively. These games continued until the central pile was empty, and everyone’s savings pile points were entered into a raffle for a pair of \$20 gift cards.³ The game play for one of these tests is shown in Figure 3. We see the players initially start with a few small groups, but one player (`cx410`) attracted multiple others and was able to protect himself with many P-cards. Gameplay in the other games was similar, with one or two “leaders” emerging quickly, who sometimes engaged in a “status battle” as they try to out-buy the other in emojis. These battles were usually resolved when one or the other hit a King or Queen. The distribution of wealth in the human games typically ended with one player with 1/4 of the wealth, about half the players with half the wealth, and the remaining half with little.

IV. DISCUSSION

A key element in group formation is the emergence of *leaders*, and therefore, also *followers*. Leaders, or “the elite” usually obtain more resources and wield the power to shape social structure, and thereby solidify their position as leaders [6]. Trust-ya encodes some of the processes that underlie this status formation process, in the hopes that it can be used as a simulation environment to better understand them.

We ground Trust-ya in the status processes succinctly formulated by Ridgeway [4], who discusses dominance hierarchies (e.g. leader-follower groups), and shows how they require perceptions of perceived competence, as well as status beliefs about this same competence of other group members. Ridgeway’s theory in some sense sums up a long sociological investigation that is referred to broadly as *expectation states theory*, broadly based on the idea of reality, and status in particular, as something both objective (related to the real world) and subjective (related to the social structure of the world for any agent) [7]. In expectation states theory, the dynamics of status follows a reflexive feedback loop. First, diffuse status characteristics are assigned based on observable status cues, such as gender, race, or physical artifacts: “...people jockey for position in the micro-dynamics of status

³These games were played using a simpler interface than that shown in Figure 1 (e.g. the images of playing cards and colors were missing), but the elements of the game were the same. To maintain a J,Q,K P-card a player needed 1, $N/4$ and $N/2$ supporters, respectively.

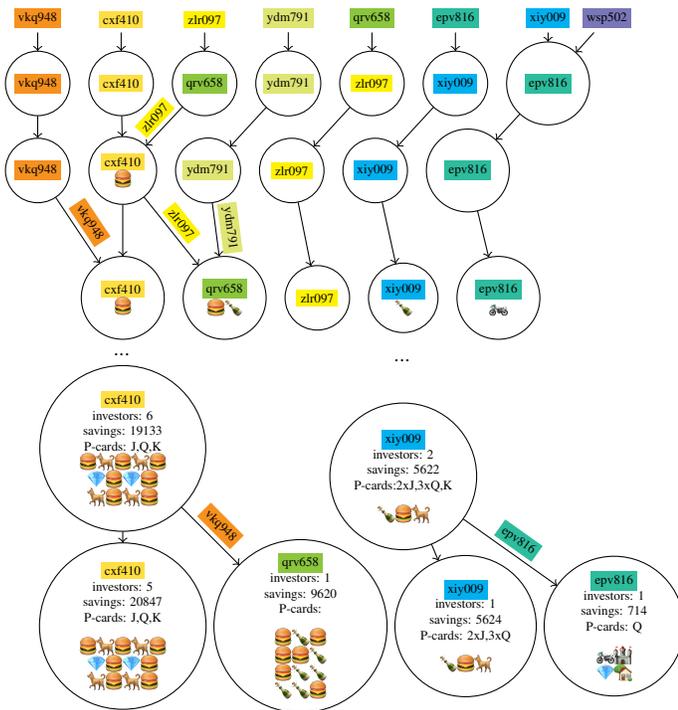


Fig. 3. Game play of an all-human game with eight players showing emoji and groups formed. Nineteen mid-game stages are omitted, in which there were no significant changes other than increased emoji purchases, and a slow migration of most investors towards `cxf410`. Nodes show the number of investors in a player, and that player’s savings pile, P-cards, and status symbols.

through objects they display” [4, p.113]. Second, these diffuse status characteristics modify expectations for performance in ways that are self-fulfilling. Agents with initially assigned high status are able, by using their increased resources, to define the direction the group takes in ways that are most conducive to maintaining the status hierarchy so defined. This status may then be challenged or found lacking in evidence by the group. Alternatively, low-status actors may be able to rise in status should they display some assertiveness, but not too much as this will be seen as disingenuous or deceptive. Leaders need to display competence, but also fairness and interest in a shared group identity with their followers [4][8].

In Trust-ya, status is observable only through each user’s displayed emoji (and their random name). However, the objective status is really defined by a player’s P-cards, which are hidden from the others (but possibly revealed during payouts). P-cards naturally reinforce an existing hierarchy in the social structure by forcing lower status groups to realize the expectations of higher status groups who control resources. The lower status group is then compelled to accept the meaning of the social relationship as defined by the higher status group, acceptance of which involves a re-definition of their position as one of deference. Higher status groups obtain more resources and thereby can better control the environment to suit their abilities, improving their status. To replicate this in Trust-ya, players with more P-cards get more resources when investing.

The correspondence continues if one considers that the display of status is something agreed upon by both low and high status actors. Although Trust-ya is currently built using a

ranked set of status symbols, these symbols may be stripped of meaning by offering them at an equal price. Meaning can then emerge as each symbol is eventually reified by the group through gameplay. Over the course of our test games, we observed this process occurring for the hamburger (cheapest emoji). One player, Bob, bought many hamburger emoji in the first game, and ended up winning the game. Thereafter, the hamburger emoji became associated with winning and with Bob. In a game that supported dynamic, open market pricing, we speculate that our players may have been willing to pay more for the hamburger as a symbol of Bob-ness and of winning-ness, but this conjecture has yet to be demonstrated.

There are ethical concerns with the use of data from a game such as this to infer social status and dynamics for purposes outside of research. These concerns are not addressed in this paper, but we stress that the investigation of such issues is part of our ongoing efforts and plays a critical role in the development of Trust-ya [9].

V. CONCLUSION

This paper reviews the design of a game for investigating status processes in small groups. Our objective was to create a game that would replicate some elements of status processes, including the establishment of leader-follower behaviours. To do so, the game made use of a standard social dilemma, in which one agent must defer to another to achieve jointly optimal rewards. Coupled with a game that places one player in a position of dominance or control (the sharing mechanism), this naturally lead to the expected status orders being established in groups of players. We investigated the emergence of such orders in simulated agents and in humans, and we discussed theoretical connections to social psychology. Our aim is to further develop the game and its automated agents, and to observe their influence on the behaviour of human players.

ACKNOWLEDGMENT

We thank the CHIL researchers who assisted with design iterations, and gratefully acknowledge support from AGEWELL NCE Inc., the Alzheimer’s Association, and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- [1] R. B. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, Massachusetts: Harvard University Press, 1991.
- [2] C. Tilly, *Durable Inequality*. Berkeley, CA: Univ. of Calif. Press, 1998.
- [3] R. S. Burt, *Structural Holes: The Social Structure of Competition*. Harvard University Press, 1992.
- [4] C. L. Ridgeway, *Status: Why is it everywhere? Why does it matter?* Russell Sage, 2019.
- [5] C. Gini, “On the measure of concentration with special reference to income and statistics,” *Colorado College Publishers*, vol. 208, no. 1, pp. 73–79, 1936.
- [6] T. Piketti, *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press, 2014.
- [7] P. L. Berger and T. Luckmann, *The Social Construction of Reality*. Doubleday: Anchor Books, 1967.
- [8] N. K. Steffens, S. A. Haslam, K. Peters, and J. Quiggin, “Identity economics meets identity leadership; exploring the consequences of elevated CEO pay,” *The Leadership Quarterly*, vol. 31, 2020.
- [9] L. Stark and J. Hoey, “The ethics of emotion in AI systems,” in *Proc. FaCCT**. ACM, 2020, pp. 782–793.