# The Colored Trails Framework

## Modeling Human Negotiation in Strategic Settings

Sevan G. Ficici, Ya'akov Gal, Avi Pfeffer

*Division of Engineering and Applied Sciences*

Harvard University

## One-Shot Three-Player Game

### Scenario

- Two proposers ('allocators'), single responder ('deliberator').
- Deliberator has full information.
- Allocators lack information of each other's state.
- Each allocator offers a chip exchange (or declines to exchange).
- Deliberator selects from received offers or rejects all offers.
- We model human play, and build a computer allocator agent that maximizes its expected score.

### Challenges

- Allocator Dilemmas: Allocators compete for deliberator's choice, but they do not know how good their offers need to be. (Each allocator lacks information about the other's state.)
- Deliberator Dilemmas: Allocators' offers may place deliberator's social factors in opposition. Example:

  A1 Exchange: $A1^-$ (+100 pts) $D^+$ (+5 pts)

  A2 Exchange: $A2^+$ (+10 pts) $D^+$(+10 pts)

### Social Factors

- *Individual Benefit* (IB): The gain a player obtains from a trade.
- *Pairwise Benefit* (PB): Sum of individual benefits of the players involved in a trade.
- *Simultaneous Improvement* (SI): Indicates whether both players involved in a trade have IB > 0.
- *Unfairness of Division* (UD): Distance to fairest feasible trade (the trade that comes closest to evenly dividing PB).
- *Marginal Cost for Marginal Gain* (CG): The cost to one player to improve the benefit to another player vs. the benefit.

### Models and Learning

- Deliberator: Three-way choice (A1, A2, Reject Both).
- Allocators: N-way choice between possible offers to deliberator.
- Mixture model over generalized linear models (Gal & Pfeffer 2006).
- Learning method is EM interleaved with gradient descent.

(1) $f^{(t)}(c) = \alpha_1^{(t)} F_1(c) + \alpha_2^{(t)} F_2(c) + \ldots + \alpha_n^{(t)} F_n(c)$

(2) $\Pr(c \mid \mathcal{C}, t) = \left( e^{f^{(t)}(c)} \right) / \left( \sum_{c \in \mathcal{C}} e^{f^{(t)}(c)} \right)$

### Maximizing Allocator Agent

1. **INPUT:** Deliberator state $S_D$ and Allocator 1's state $S_{A1}$
2. Let $M_{A1}$ be Allocator 1's move (offer) set, given $S_{A1}$ and $S_D$
3. Let $S_{A2}(i)$ be $i$-th ($i \in [1, n]$) i.i.d. draw from r.v. $\mathcal{S}_{A2}$
4. Let $M_{A2}(i)$ be Allocator 2's move (offer) set, given $S_{A2}(i)$ and $S_D$
5. Let $M_{A2}(i, j)$ be Allocator 2's $j$-th ($j \in [1,m]$) possible move (offer)
6. Let $\rho_{A2}(i, j)$ be Pr(Allocator 2 chooses $M_{A2}(i, j) \mid S_{A2}(i), S_D$)

   $\rho_{A2}(i, j) = \sum_{u \in T(A)} \Pr(u) \cdot \Pr(M_{A2}(i, j) \mid M_{A2}(i, -j), S_{A2}(i), S_D, u)$

7. $\hat{A} = \operatorname{argmax}_{A1 \in M_{A1}}$ — A1 move that maximizes...

   $1/n \cdot \sum_{i=1,n}$ — mean over all A2 state samples...

   $\sum_{j=1,m}$ — over all A2 moves for this state...

   $[\ \rho_{A2}(i, j) \cdot$ — Pr(A2 chooses this move) • ...

   $\sum_{t \in T(D)} \cdot \Pr(t) \cdot$ — mean over all D types of... expected payoff to A1:

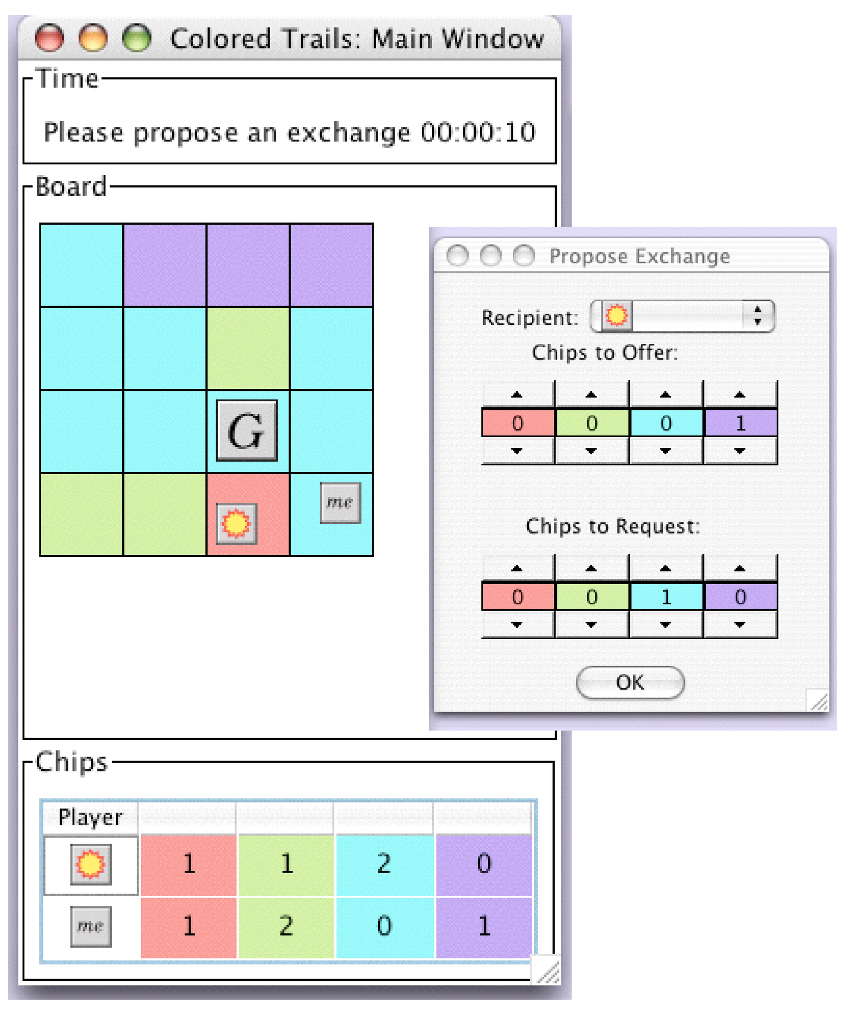
   $\sum_{c \in C} \Pr(c \mid A1, M_{A2}(i, j), \langle S_{A1}, S_{A2}(i), S_D \rangle, t) \cdot \pi(A1, c)\ ]$

8. **OUTPUT:** $\hat{A}$

## What is Colored Trails ?

### Description

- The Colored Trails (CT) framework is a test-bed for modeling and learning decision-making in social contexts.
- CT is a board game that can be played by people, computers, or heterogeneous groups.
- The game board is a grid of colored squares. Each player has a piece on the board, a designated goal square, and a set of colored chips (resources) it can use to reach its goal.
- To move its piece to a neighboring square, a player must have a chip of the same color as the square.
- When one or more players lack the resources needed to reach the goal, they may negotiate chip exchanges.
- A player's score may depend upon her own location, remaining resources, and/or those of other players.
- The game's complexity can be varied by changing the board layout, chip distributions, and scoring function.
- CT provides an analogue for task settings in multi-agent systems: chips represent resources; goal squares represent objectives; players achieve objectives by exhausting resources as they move across the board.




Snapshot of the CT GUI interface for a 2 player game. The "me" player needs a cyan chip to get to goal while "sun" can independently get to the goal. The "me" player has offered an exchange of 1 purple chip in return for a 1 cyan chip.

### Roadmap

- This poster outlines two of several on-going research projects currently underway at Harvard University.
- Other studies include
  - An analysis of the influence of players' personal relationships on bargaining behavior (Marzo *et al.* 2004)
  - A decision-making model that conditions agents' helpfulness on their estimate of others' (Talman *et al.* 2005)
- Initial development of CT is supported by NSF. Further research and disseminating the framework is supported by DARPA.

### Obtaining the CT Framework

- CT will soon be released as open-source software package.
- Package consists of: Server for controlling the game, GUI clients for human play, and Java templates for creating automatic agents.
- Package also includes a language for configuring the complexity of the game along different parameters.

## Iterated Two-Player Game

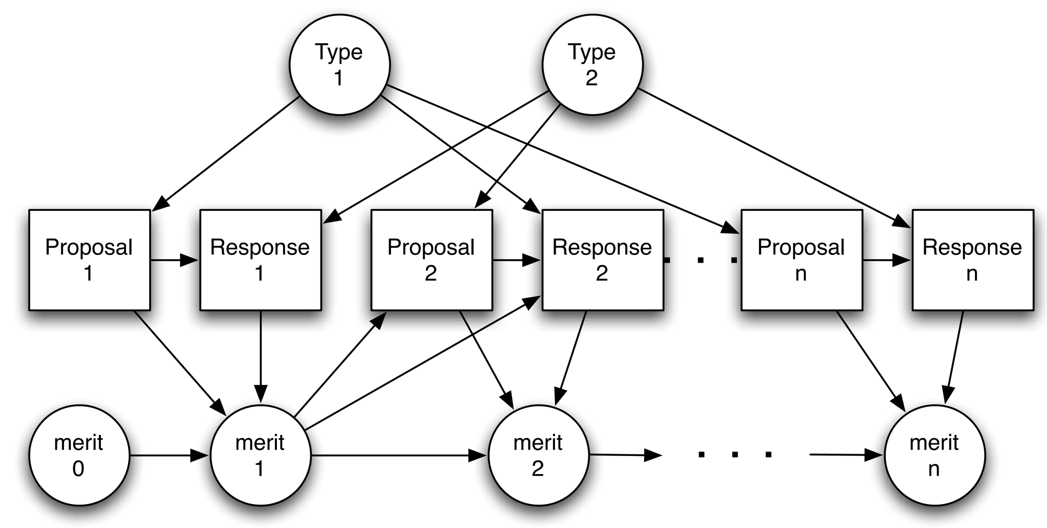
### Scenario

- Two players play a series of take-it-or-leave-it CT rounds.
- Each round differs in board layout, players' roles, chip distribution and dependency relationships between players.
- Collected data includes 57 games, each played by different pairs of people. Each game consists of between 2 to 8 take-it-or-leave it CT rounds. Proposer and responder roles switch at every round.
- Objective is to build a computer player that could outperform both humans and game theoretic computer players in varying game contexts.

### Challenges

- Because players' reciprocate each others' actions, need to
  - model how actions affect their reputation.
  - reason about the future consequences of players' actions.
- Must predict strategies for both proposers and responders because roles are switched at each game.
- Each round of play is *not* independently identically distributed; reasoning about players' offers in future rounds affects their strategies in current rounds.

### Multi-agent Diagram of *1* Game (*n* rounds)




### Proposed Model

- Each players' reputation is represented as a "merit" scalar with range [-1, 1].
  - The "merit" is computed as the difference between players' material benefit and their entitled benefit, given by the Nash Bargaining equilibrium strategy.
  - CT rounds are conditionally independent of each other given players' merits and their types.
- A player's utility for a potential action is a weighted summation of
  - social preferences of potential action.
  - other player's "merit".
  - future ramification of potential action by sampling games from the future.
- Use a mixture model of types. Each type has a separate weight distribution for social preferences.

### Computing Ramification of Action