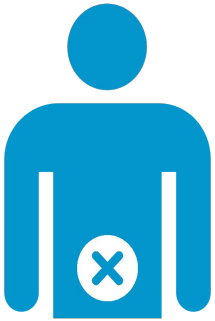
# Aligning with Heterogeneous Preferences for Kidney Exchange

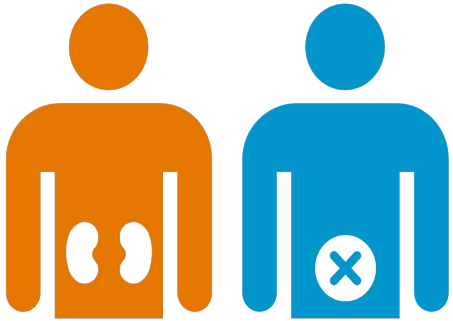
By Rachel Freedman
Center for Human-Compatible AI (CHAI)
UC Berkeley

# Kidney Exchanges
# Preference Aggregation
# Methodology and Simulations

# Kidney Exchanges

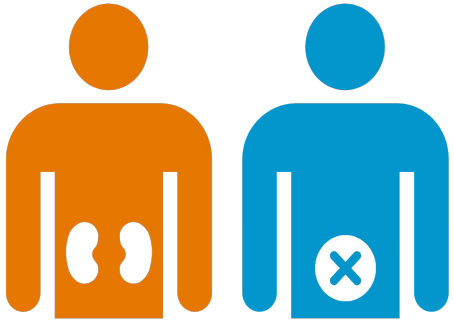
## Preference Aggregation Methodology and Simulations

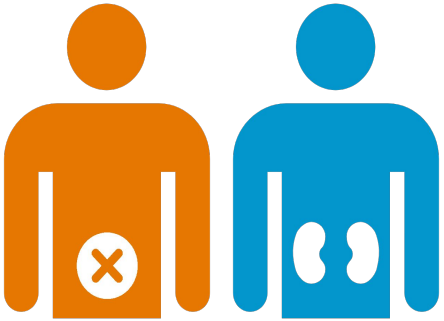




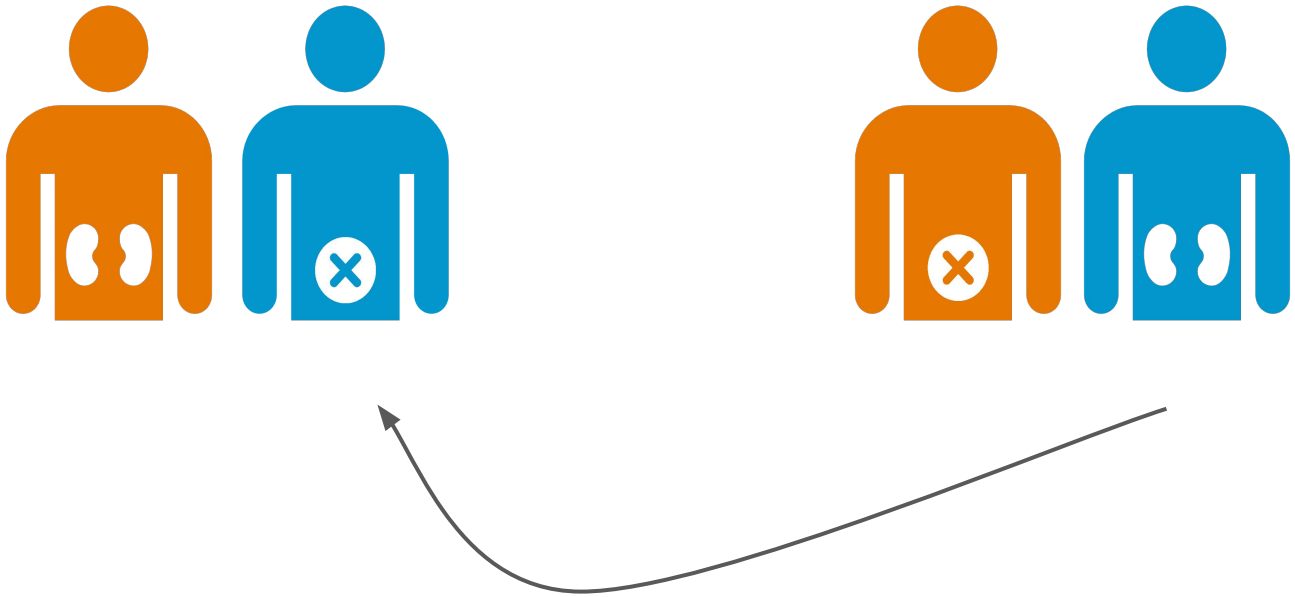
Rachel Freedman, AISafety 2020

Rachel Freedman, AISafety 2020

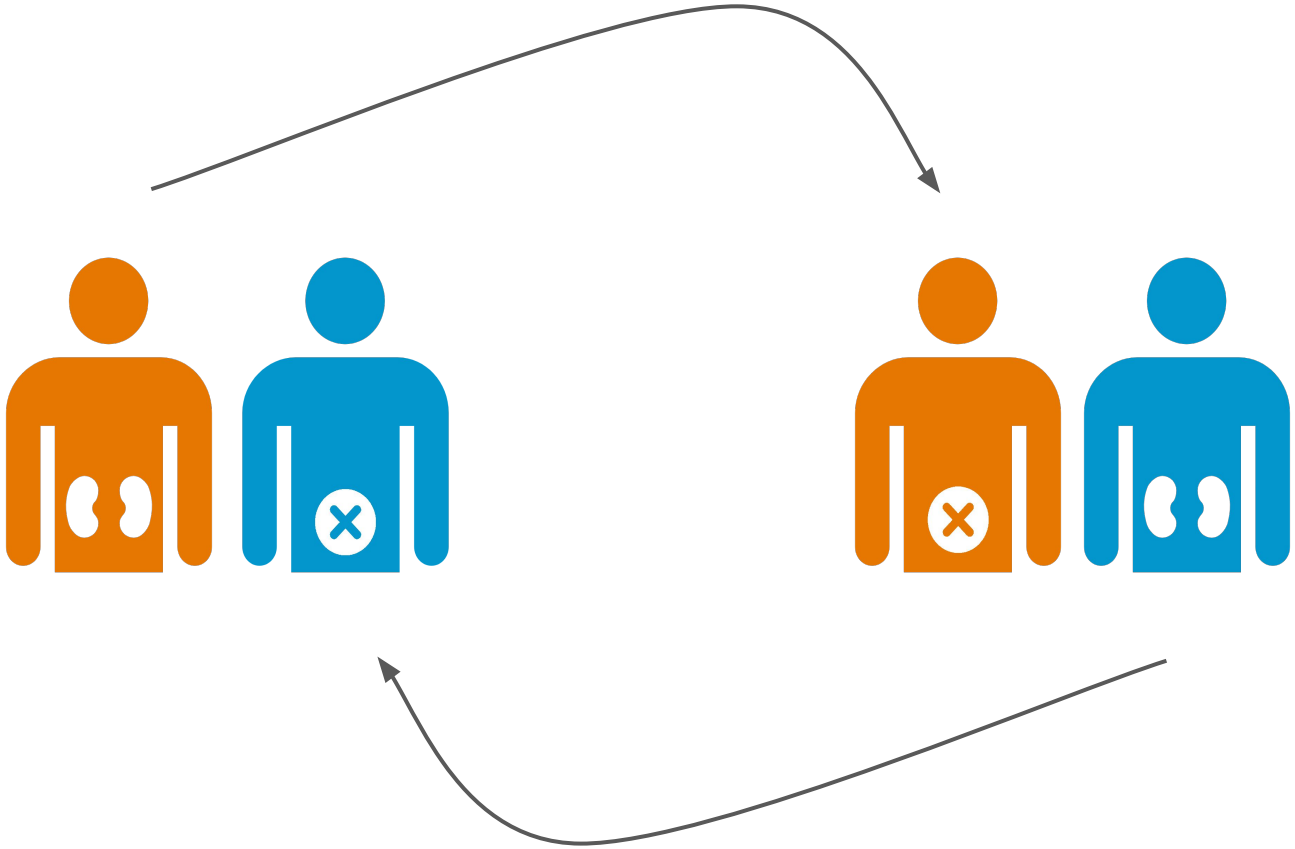




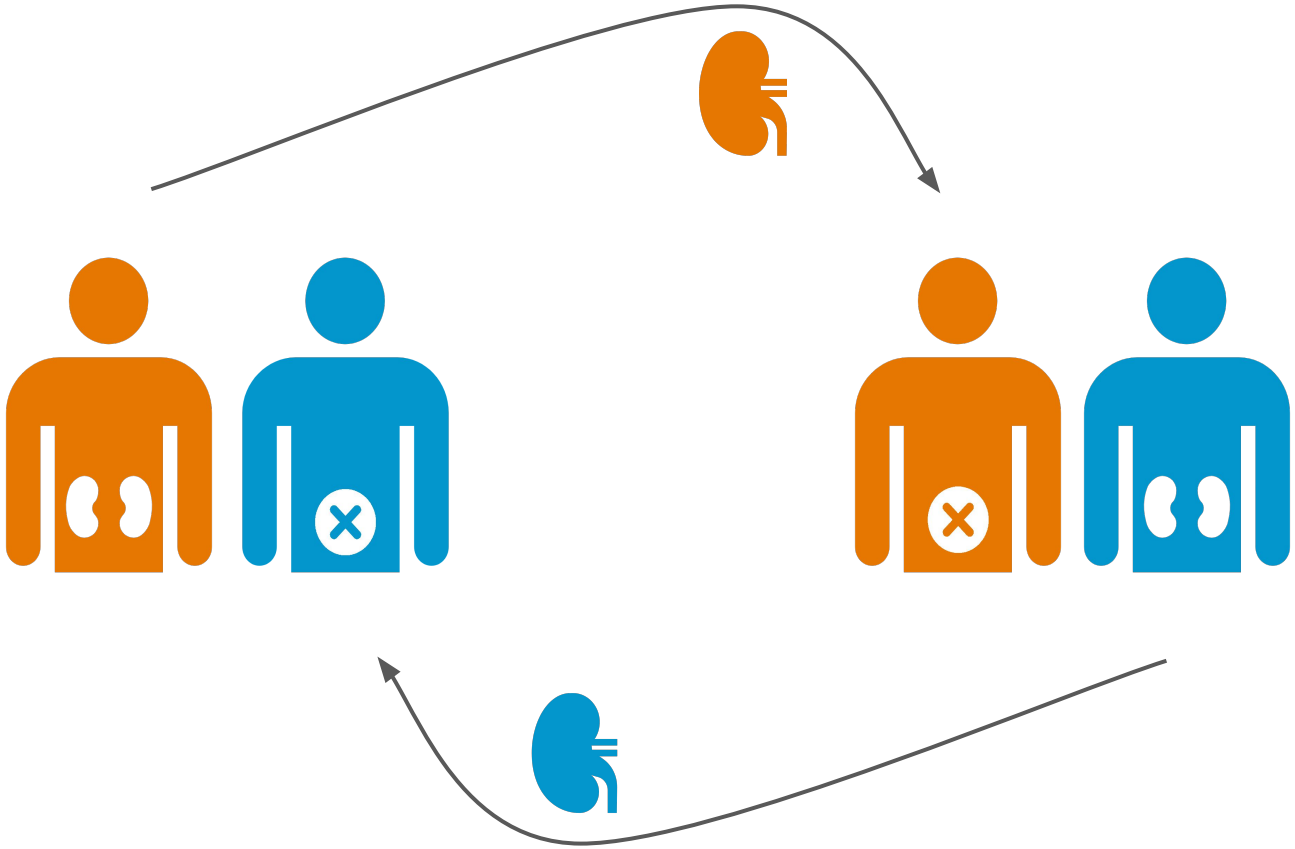
Rachel Freedman, AISafety 2020



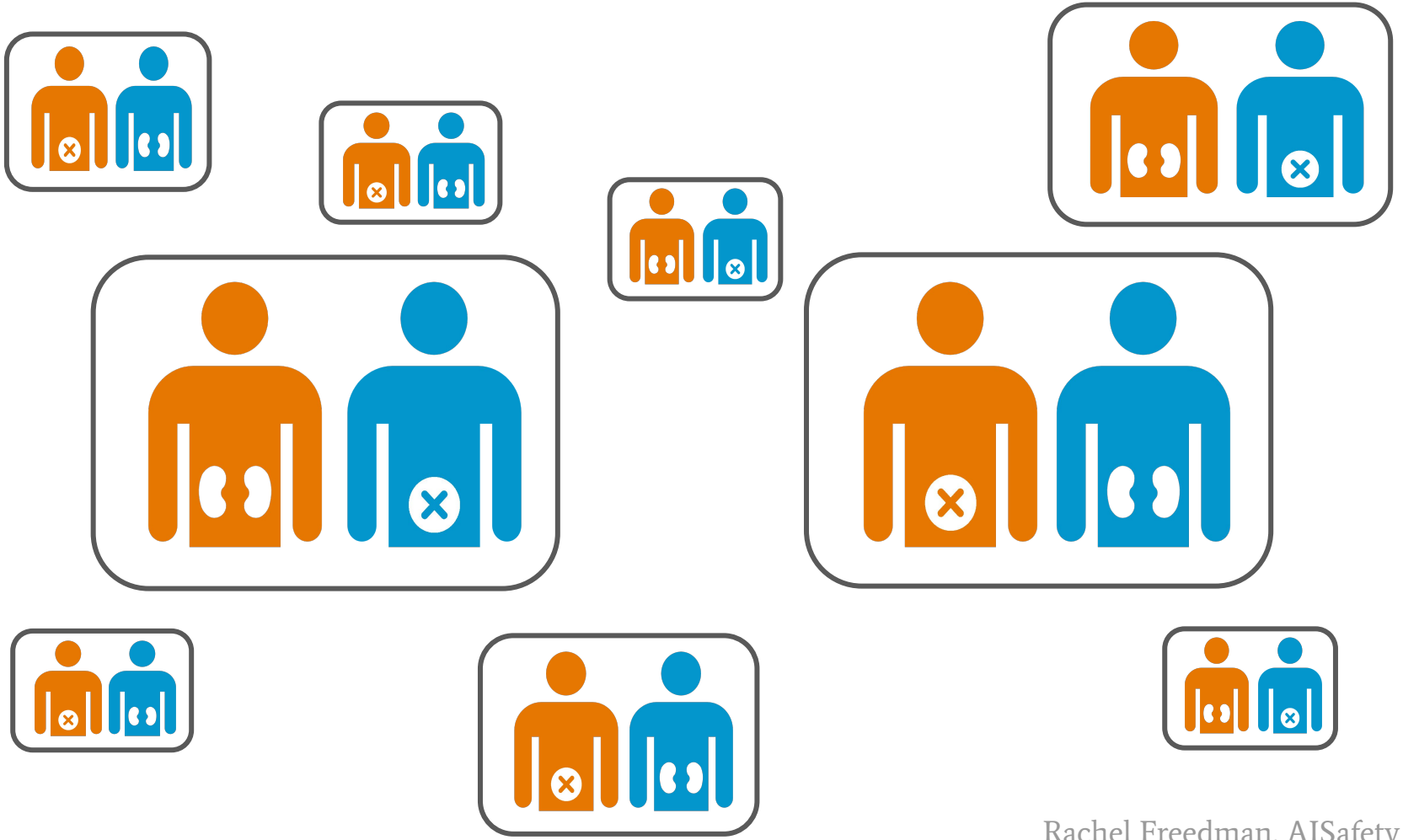
Rachel Freedman, AISafety 2020



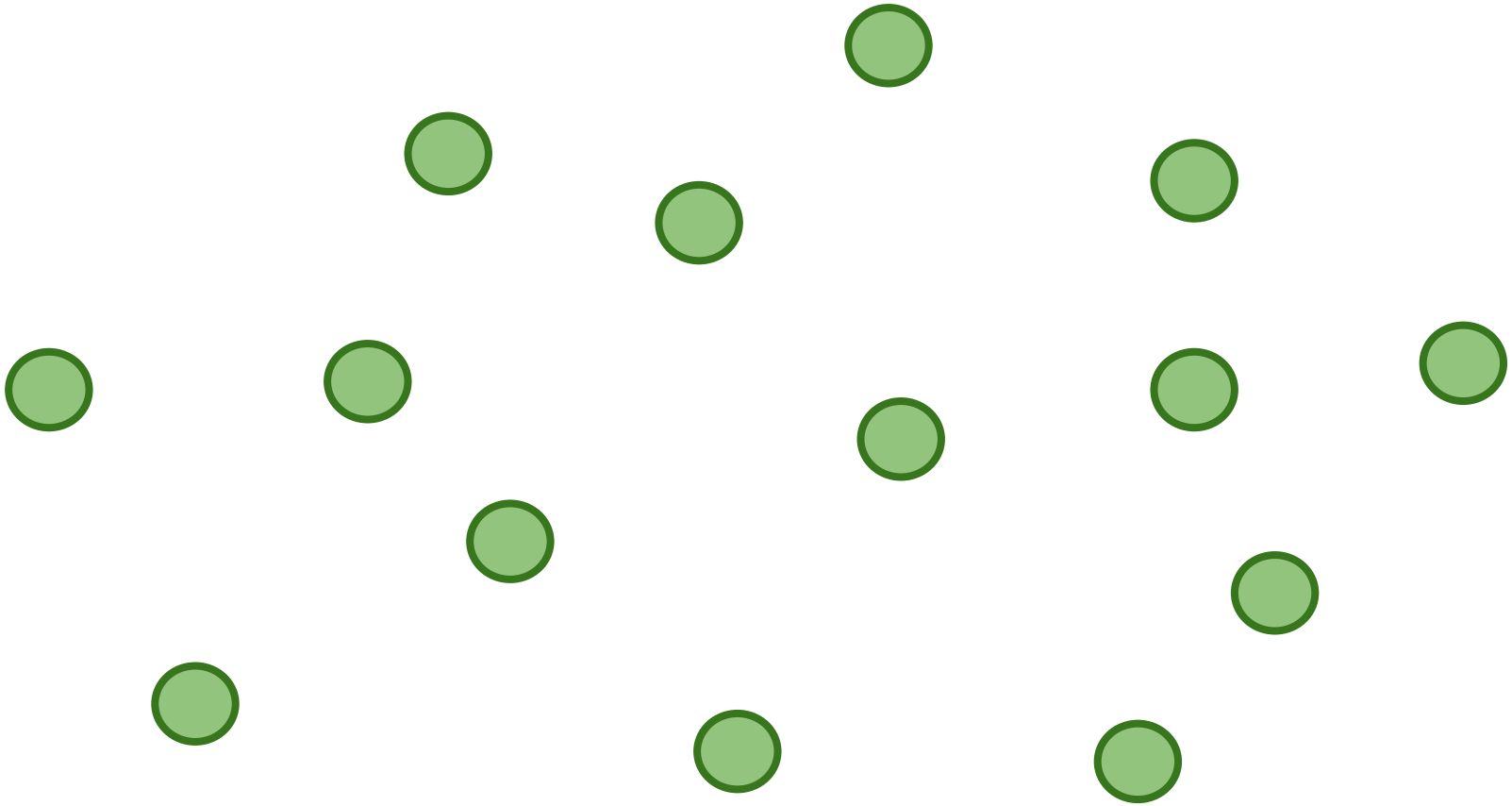
Rachel Freedman, AISafety 2020



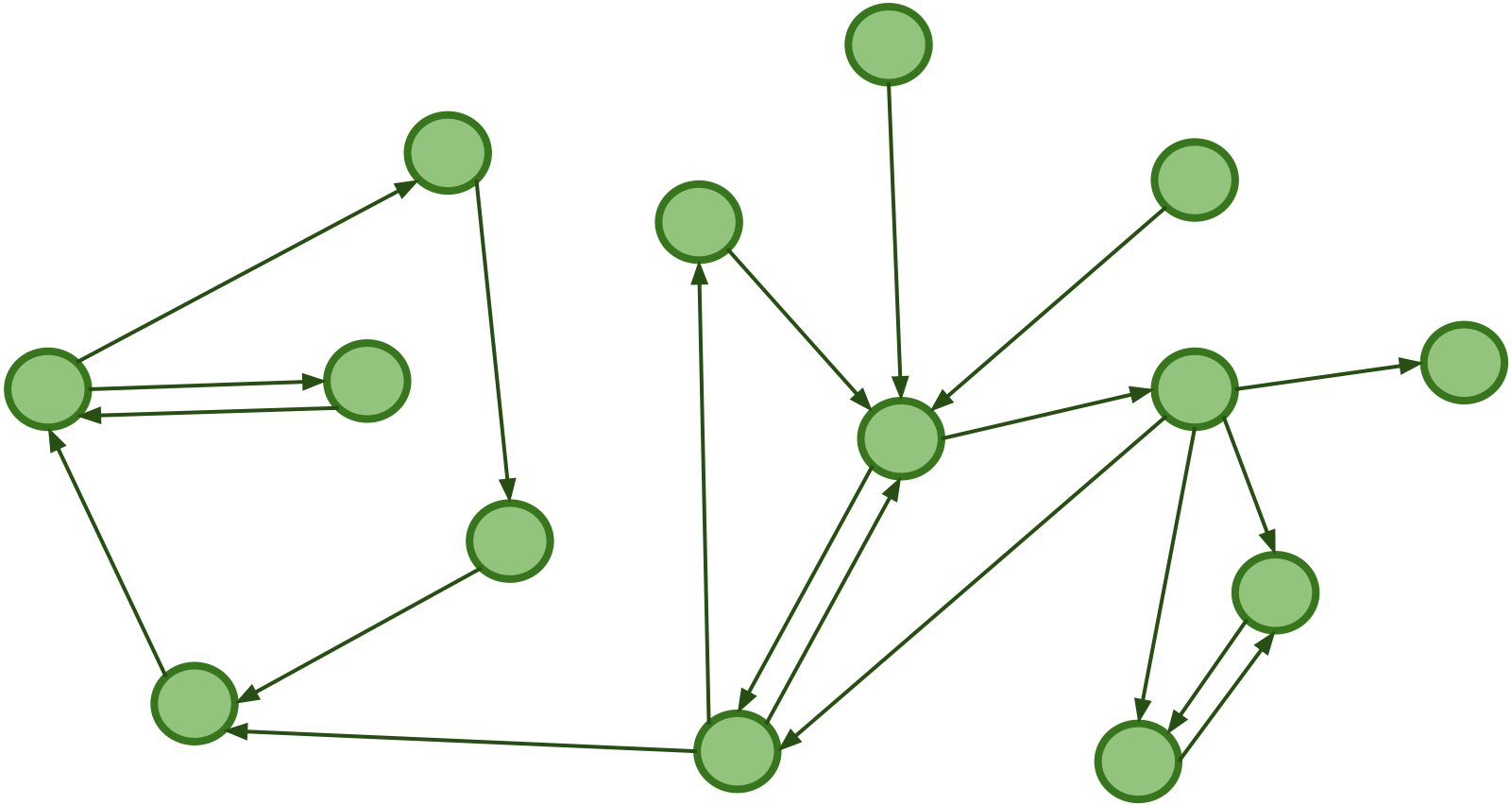
Rachel Freedman, AISafety 2020



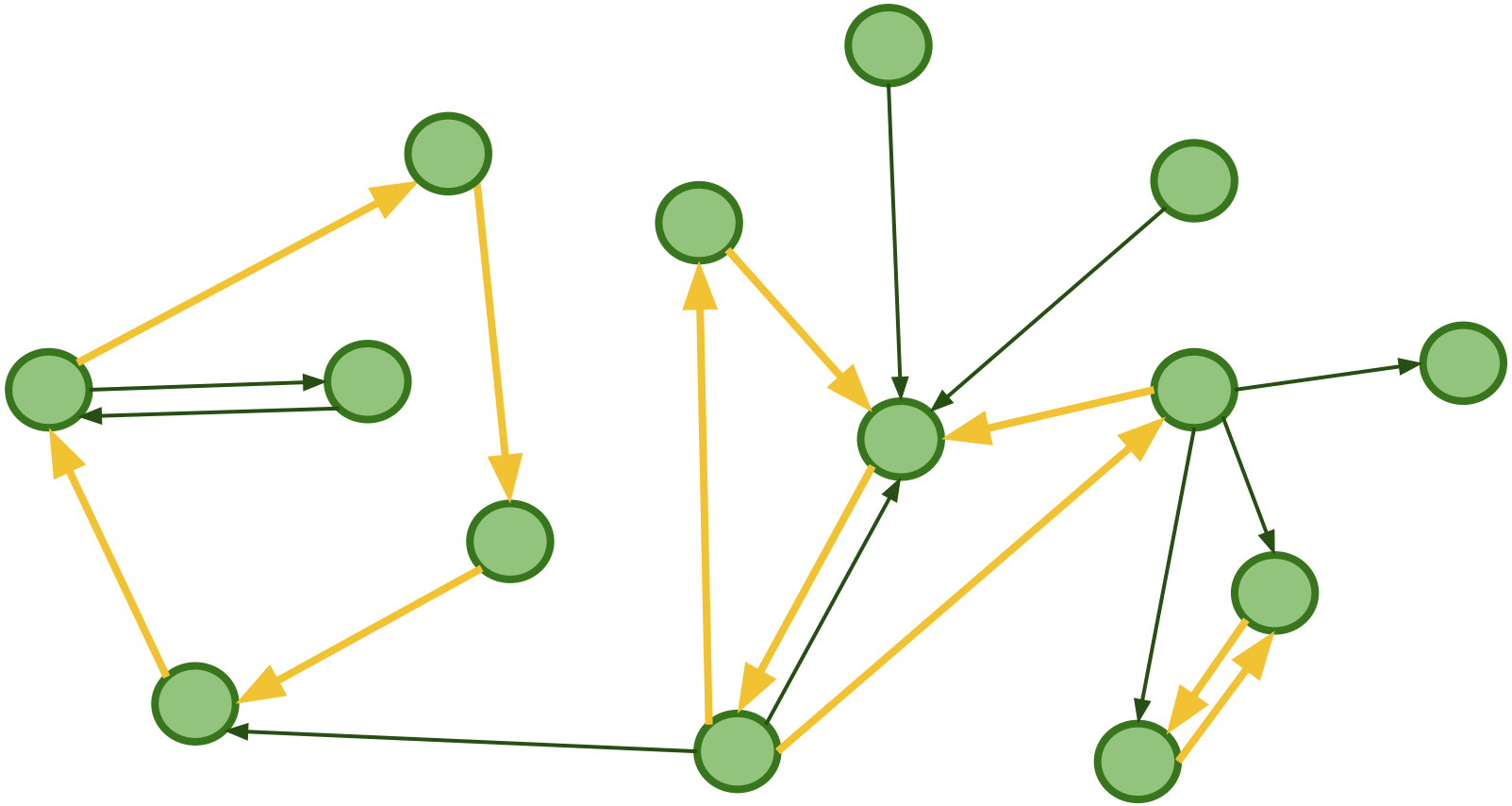
Rachel Freedman, AISafety 2020



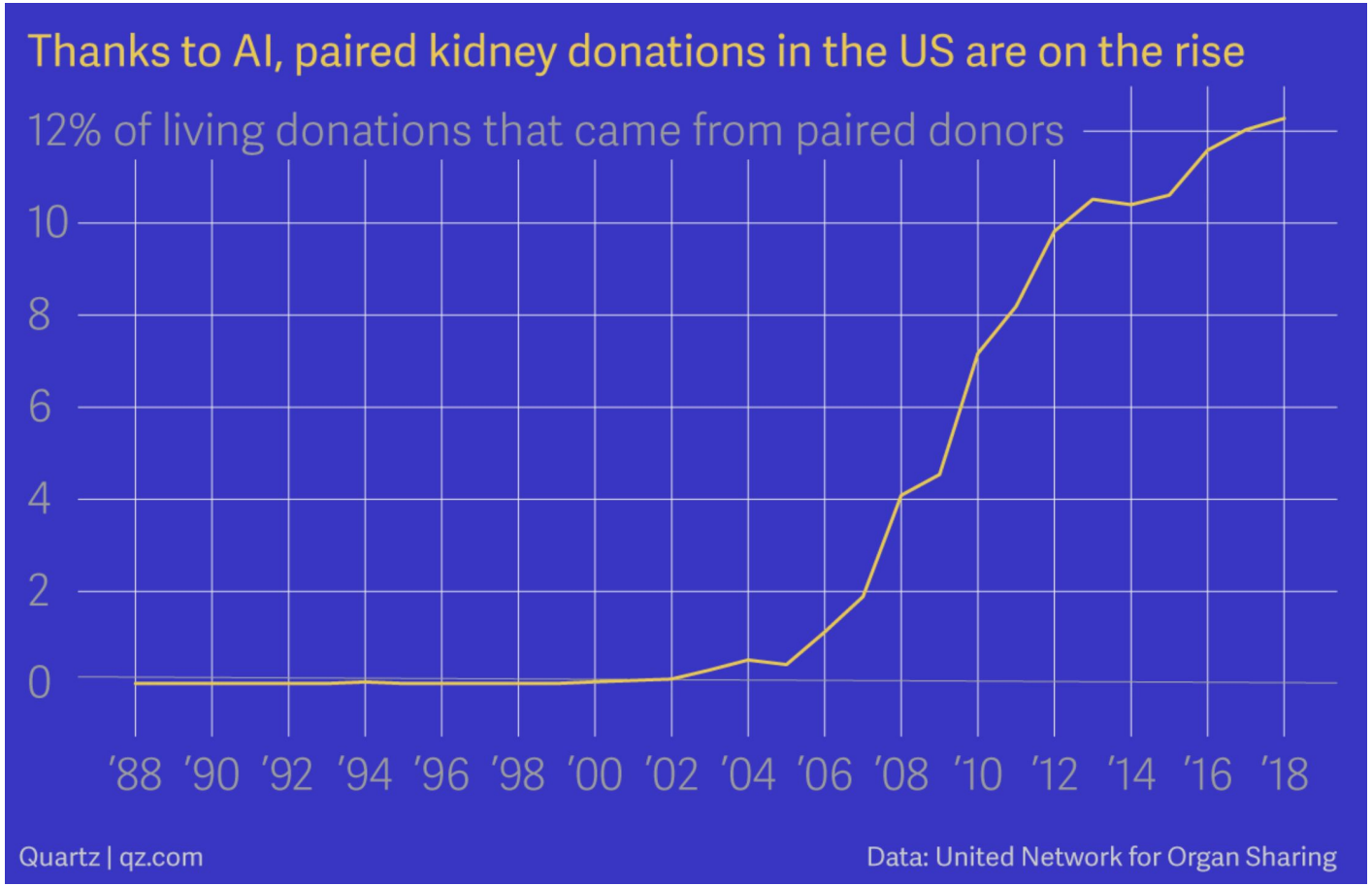
Rachel Freedman, AISafety 2020



Rachel Freedman, AISafety 2020

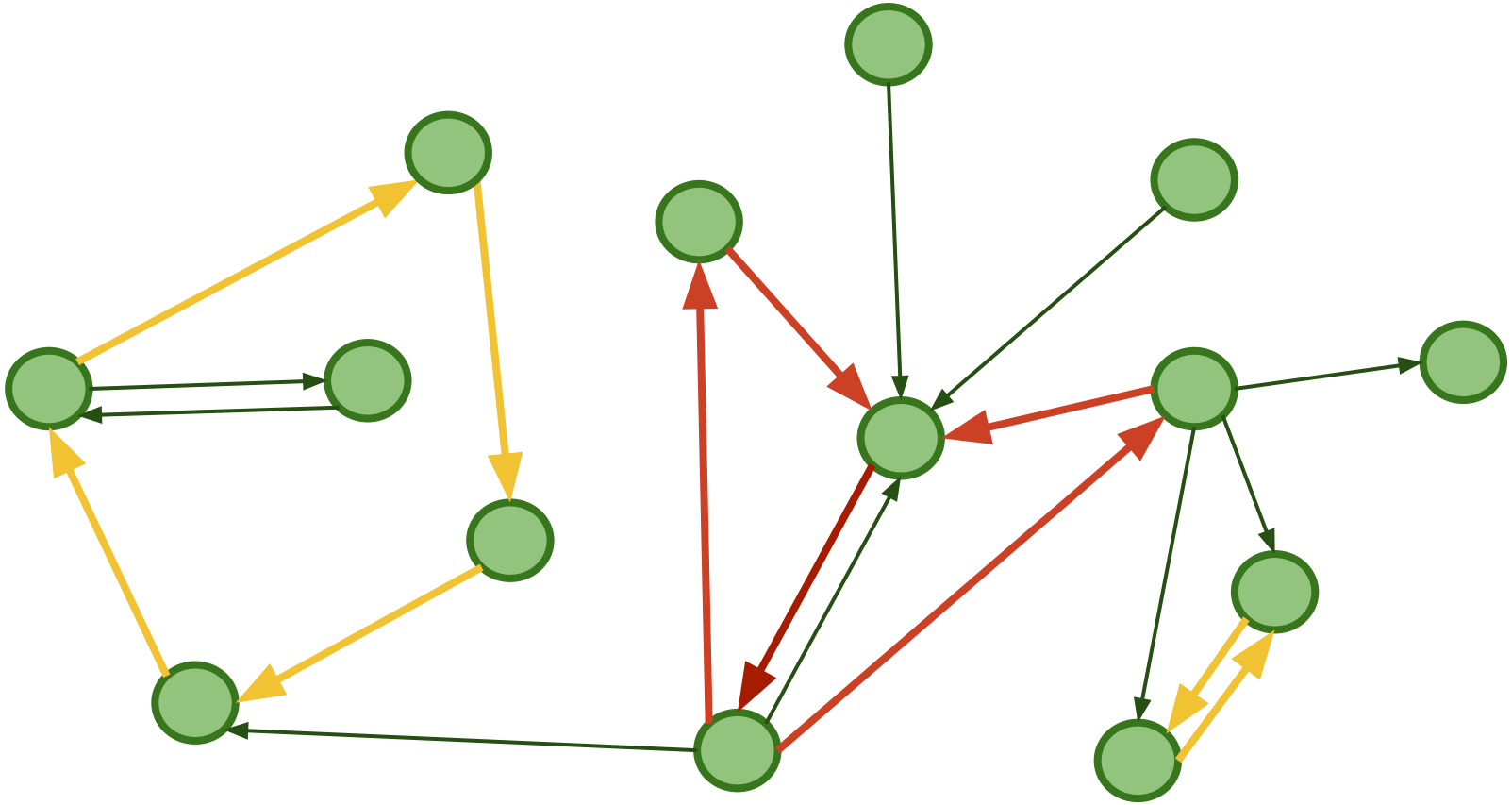


Rachel Freedman, AISafety 2020




Source: Quartz, *How AI changed organ donation in the US*

Rachel Freedman, AISafety 2020



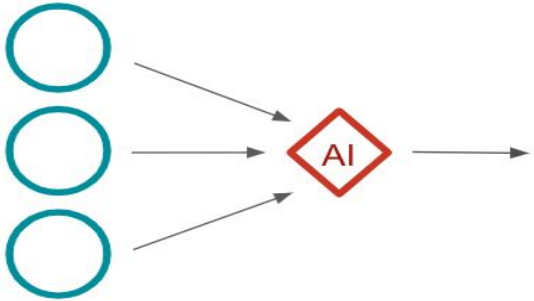
Rachel Freedman, AISafety 2020

Kidney Exchanges

# Preference Aggregation

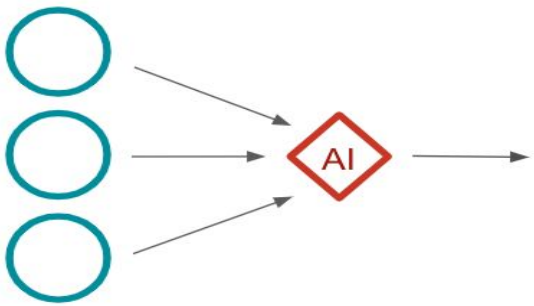
Methodology and Simulations

# Value Alignment as Preference Aggregation



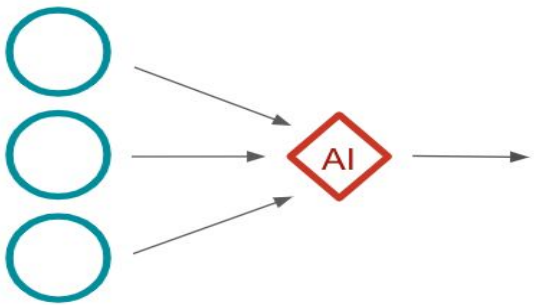

Rachel Freedman, AISafety 2020

# Value Alignment as Preference Aggregation




- **"Multi-single Delegation"** from *ARCHES: AI Research Considerations for Human Existential Safety* (Andrew Critch and David Krueger, 2020)

# Value Alignment as Preference Aggregation




- **"Multi-single Delegation"** from *ARCHES: AI Research Considerations for Human Existential Safety* (Andrew Critch and David Krueger, 2020)
- Examples:
  - **kidney allocation** (ex. *Adapting a Kidney Exchange Algorithm to Align with Human Values*, Freedman et al. 2020)
  - **self-driving cars** (ex. *The Social Dilemma of Autonomous Vehicles*, Bonnefon et al. 2016)
  - **bail decisions** (ex. *Human Decisions and Machine Predictions*, Kleinberg et al. 2018)

# Individual Preferences: Pairwise Comparisons

Patient A is **70 years old**, has **1 alcoholic drink per month**, and has **no other major health problems**.

Patient B is **30 years old**, has **5 alcoholic drinks per day**, and has **skin cancer in remission**.

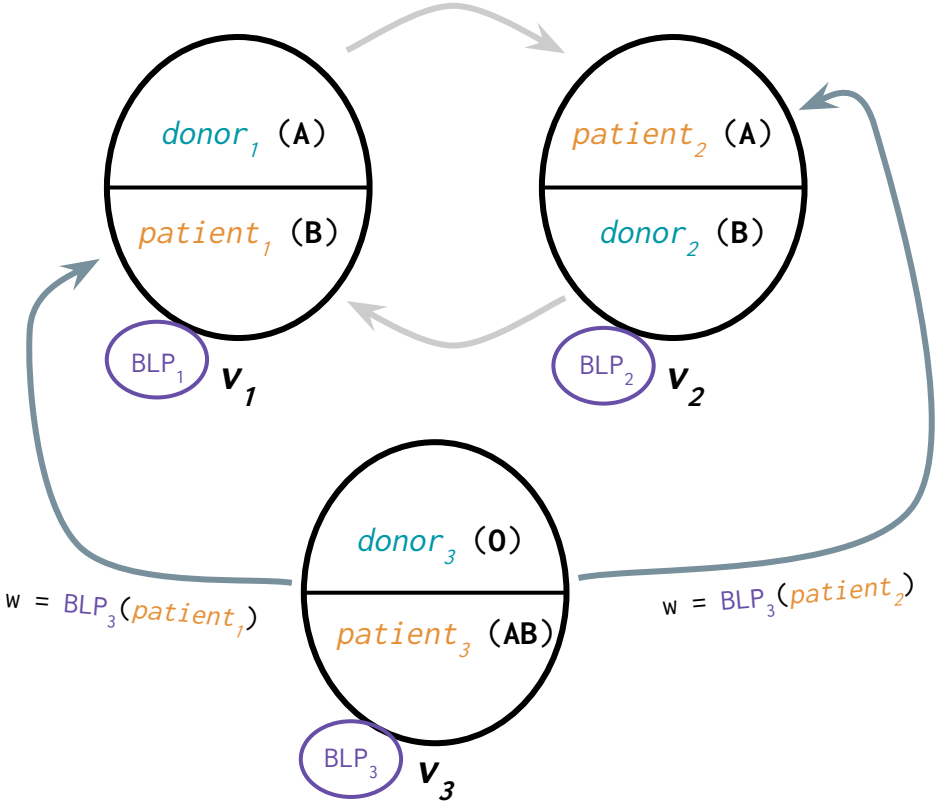
Who should get the kidney?

*Survey data graciously shared by the Duke Moral AI team.

Rachel Freedman, AISafety 2020

# Kidney Exchanges
# Preference Aggregation
# Methodology and Simulations

# Model: BLP*




*From Berry, Levinsohn, and Pakes (1995, *Econometrica*)

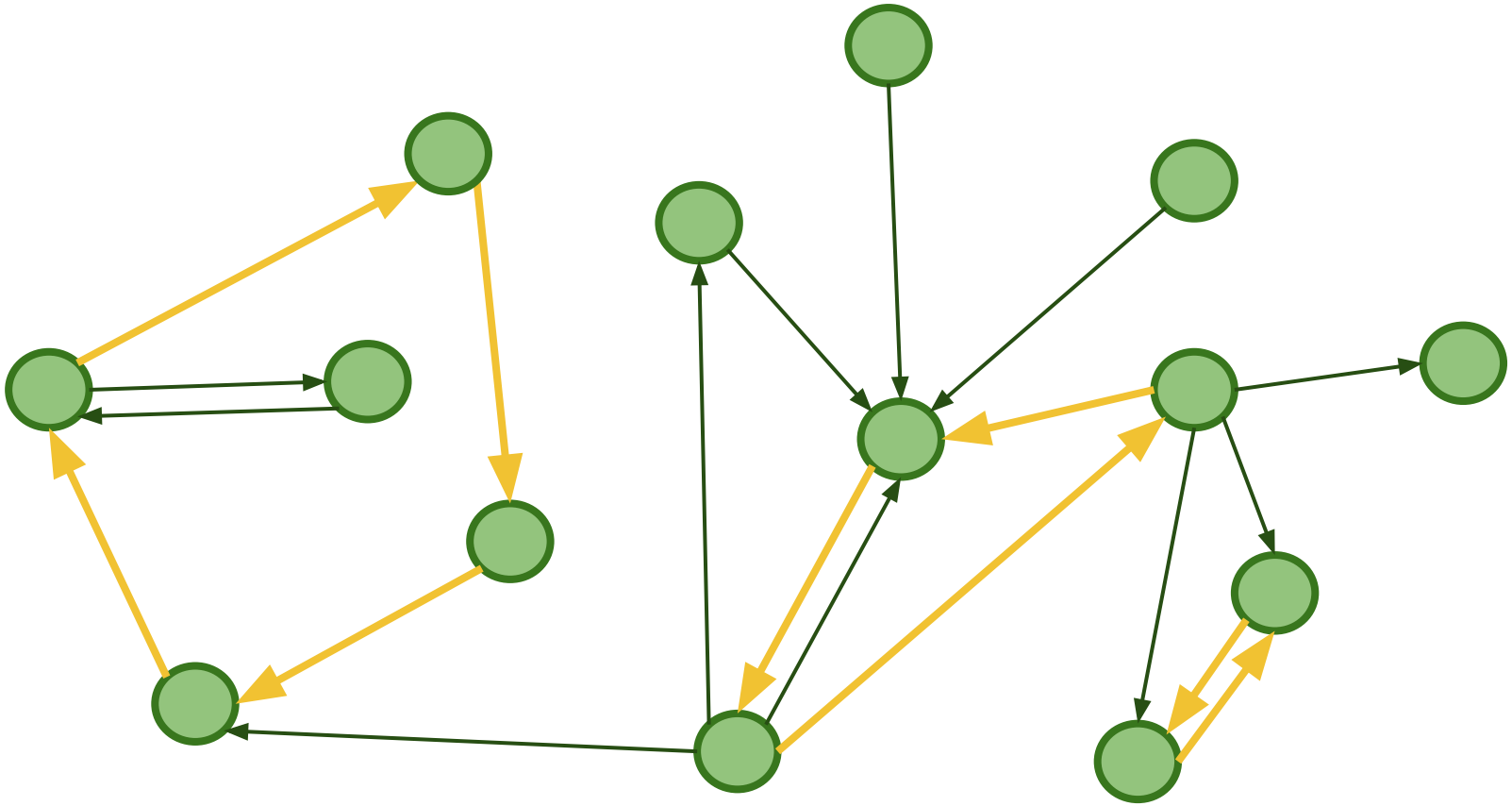
# Algorithm: Modified LP

$$\max \sum_{c \in C(L)} x_c \qquad s.t. \qquad \sum_{c:v \in c} x_c \leq 1 \quad \forall v \in V.$$

**Our Modification:**

$$\begin{array}{ll} \max & \sum_{c \in C(L)} \left[ \sum_{(u,v) \in c} w_{BLP(u,v)} \right] x_c \\ \text{s.t.} & \sum_{c:v \in c} x_c \leq 1 \qquad \forall v \in V \\ & \sum_{c \in C(L)} |c| x_c \geq Q \end{array}$$

Rachel Freedman, AISafety 2020



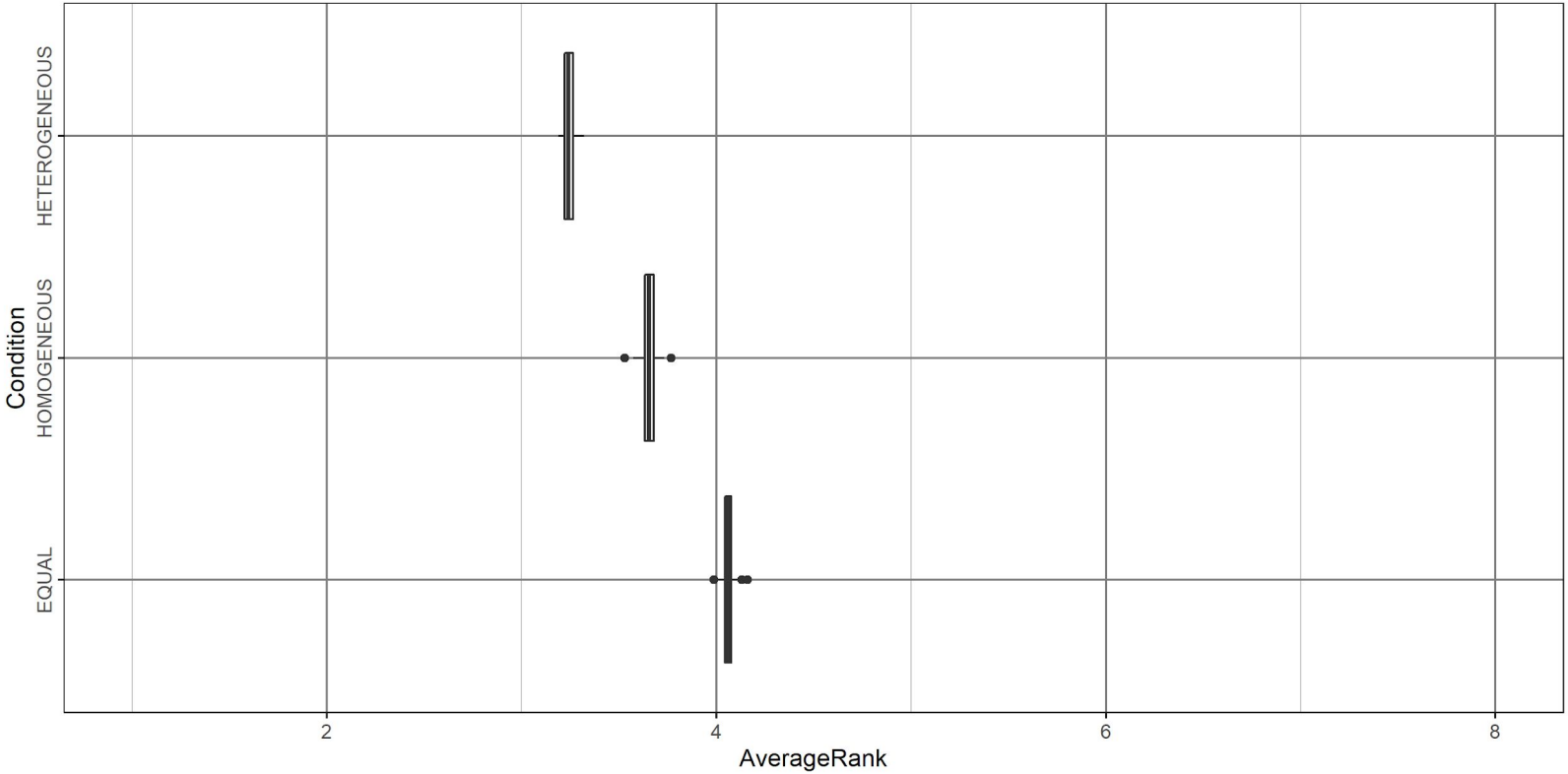
Rachel Freedman, AISafety 2020

# I built a simulation...

```java
public class VariationDriver {

        // Experimental conditions
        public enum Condition {
                EQUAL_WEIGHTS,                  // All edges have weight 1.0
                PATIENT_WEIGHTS,                // Edge weights depend on receiving patient only (BT model)
                DONOR_PATIENT_WEIGHTS           // Edge weights depend on donor and receiving patient (BLP model)
        }

        // Probabilities generated based on a match frequency of 1 day
        static final int CYCLE_CAP = 3;
        static final int EXPECTED_PAIRS = 4;
        static final int ITERATIONS = 365*5;
        static final int NUM_RUNS = 50;
        static final double DEATH = 0.00058072543318238116805064369 1;

        public static void runExperiments(){
                String expID = Long.toString(System.currentTimeMillis());
                runSimulationWithEqualWeights(expID);
                runSimulationWithPatientWeights(expID);
                runSimulationWithDonorPatientWeights(expID);
        }
```

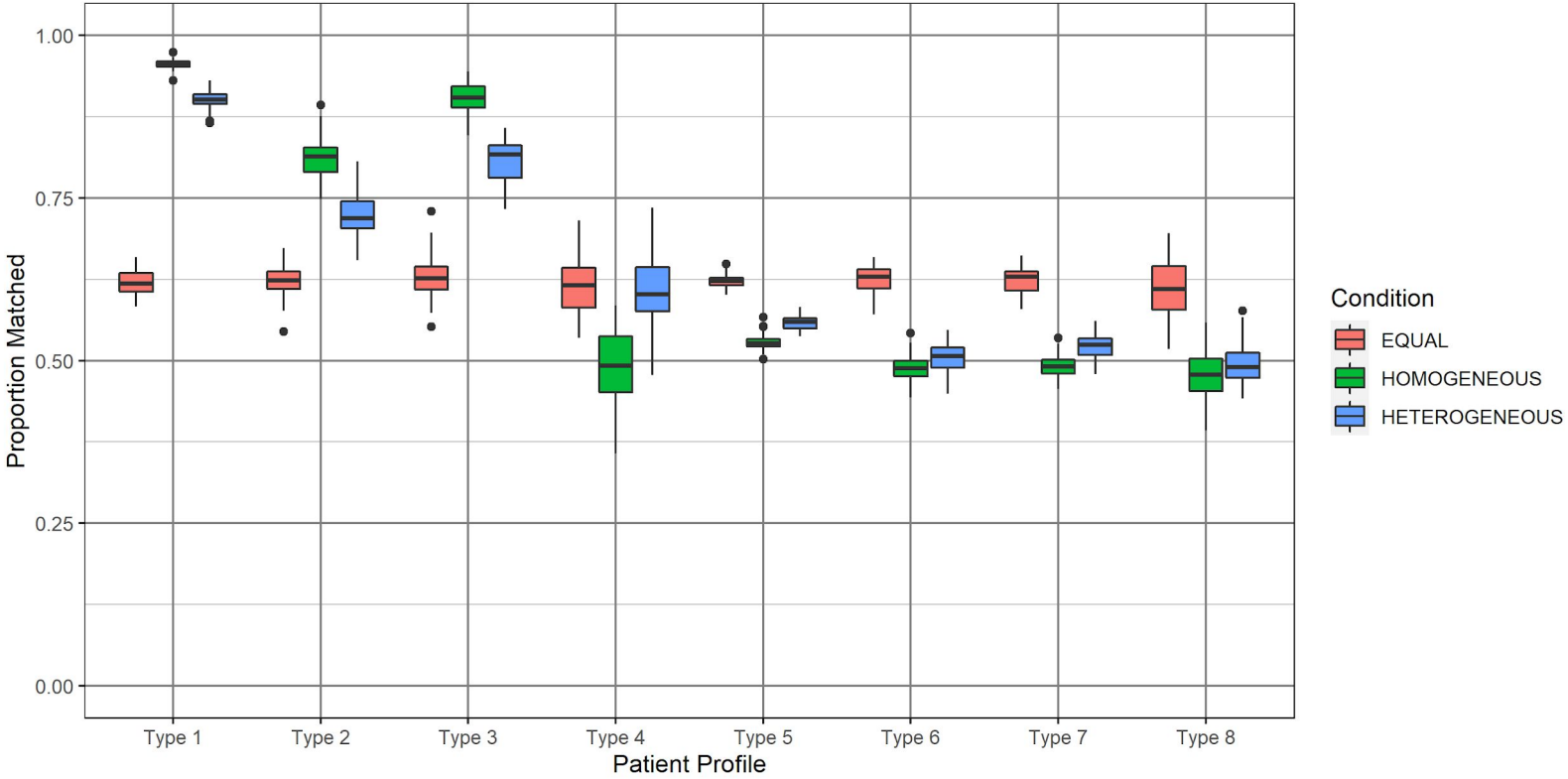
Rachel Freedman, AISafety 2020

# Considering Heterogeneity Improves Average Rankings




Rachel Freedman, AISafety 2020

# Most Desirable Profiles Still Prioritized




Rachel Freedman, AISafety 2020

# Acknowledgements

Thanks to the **Duke Moral AI team** for the human preference data.

Thanks to **Professor John Dickerson** for an earlier version of the simulation.




And thanks to **Yunhao Huang** for help on the project!

Thank you for listening!
Questions?