

# Seminar on Reinforcement Learning and Decision Making Under Uncertainty

Christos Dimitrakakis

March 24, 2023

# Outline

Expected Utility Hypothesis and the Measurability of Utility

Fairness, matching and regret

- ▶ Initially people thought decisions under risk and no risk were different.
- ▶ Can we justify choices under risk through expected utility?
- ▶ Is there an axiomatic justification?
- ▶ Can utility be measured?
- ▶ Final outcomes:  $x_1, \dots, x_n$  in  $X$ .
- ▶ Probability distributions  $F$  so that for any  $f \in F$ ,  $f_i$  is the probability of  $x_i$ .
- ▶  $f \geq g$  means 'f is not preferred to g'

The postulates are:

P1. Perfect linear order

- ▶ 1. Universality:  $f \geq h$  or  $h \geq f$ .
- ▶ 2. Transitivity: If  $f \leq g$  and  $g \leq h$  then  $f \leq h$ .

P2. Continuity


If  $af + (1 - a)h \leq g$  for all  $a \in [0, 1)$  then  $f \leq g$ .

P3. If  $a \in (0, 1)$ ,  $af + (1 - a)h \leq ag + (1 - a)h \Leftrightarrow f \leq g$ .

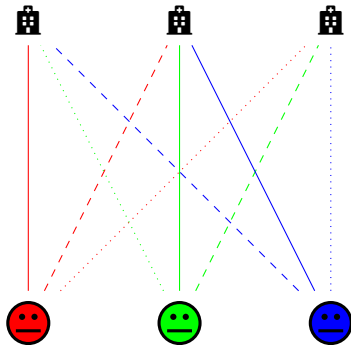
From the above postulates, the expected utility hypothesis follows.

If we assume a perfect linear order, then of course any function  $U(f)$  such that  $U(f) > U(g)$  implies  $f > g$  will be enough. Any monotonic transform of that function will also be enough. However, we can look at a special function  $C$  such that

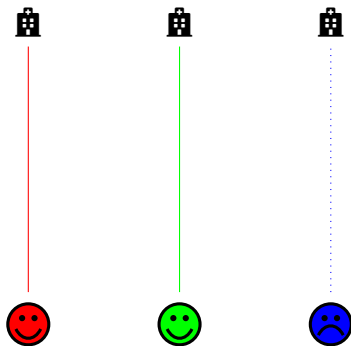
$$U(f) = E_f[C] = \sum_x C(x)P_f(x)$$

The paper does not really prove or disprove that the expected utility hypothesis. There are actually experiments which we can run that do so, but they are only obliquely referred to by this paper. 

## Matching illustration: One-sided matching



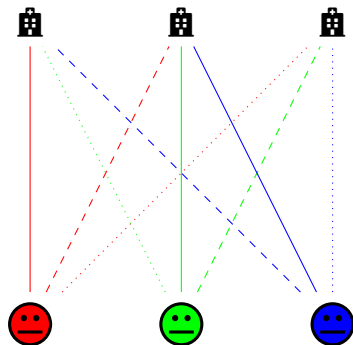
## Matching illustration: One-sided matching



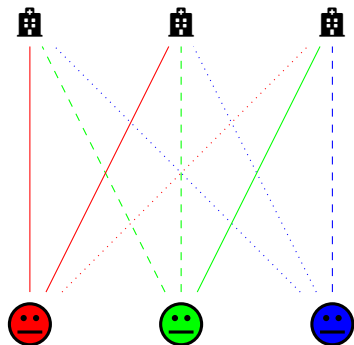
A stable match: no two agents want to swap.

# Matching illustration: Two-sided matching

Renter preferences

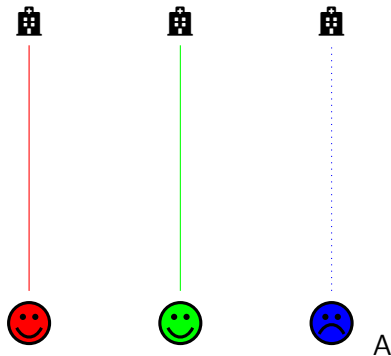


Owner preferences



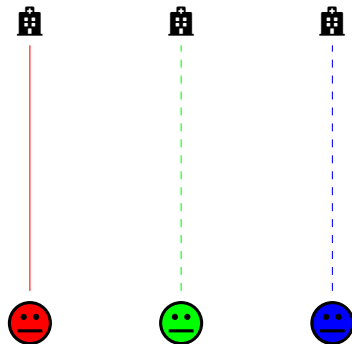
# Matching illustration: Two-sided matching

## Renter preferences



stable match: no owner and renter would like to swap!

## Owner preferences



A stable match: no owner and renter would like to swap!



# Matching problems: definition

- ▶ Agents  $i \in A, j \in B$ .

## Ordinal preferences

- ▶ Each agent  $i \in A$  has a preference  $k > j$  for  $k, j \in B$ .
- ▶ Each agent  $j \in B$  has a preference  $l > i$  for  $l, i \in A$ .

## Cardinal preferences

- ▶ Each agent  $i \in A$  prefers  $k$  over  $j$  if  $v_i(k) > v_i(j)$ .
- ▶ Each agent  $j \in B$  prefers  $l$  over  $i$  if  $u_j(l) > u_j(i)$ .

## Bi-partite Matching

- ▶  $M_{i,j} = 1$  if two agents are matched.
- ▶  $M_{i,j} = 0$  if two agents are not matched.
- ▶  $\sum_j M_{i,j} = 1$  for all  $i \in A$ .
- ▶  $\sum_i M_{i,j} = 1$  for all  $j \in B$ .

# Properties of matchings

## Stability

A matching  $M$  is stable if there is no pair  $i, j$  for which

$$M_{i,j} v_i < v_i(j), \quad M_{i,j}^\top u_j < u_j(i).$$

## Social welfare

The matchings maximising renter or owner side are stable:

$$\sum_i M_{i,j} v_i(j), \quad \sum_j M_{i,j} u_j(i)$$

## Truthfulness

If we are maximising renter's welfare, then they have no incentive to lie about their preferences.

# Learning to match

## The interaction

- ▶ At step  $t$ .
- ▶ The agents report their preferences  $N_t$ .
- ▶ The platform selects a match  $M_t$  for all agents.
- ▶ The players obtain a reward  $X_t(a, M(a, \nu_t))$
- ▶ Agents incur costs  $C$  and receive transfers  $T$

## The agents

- ▶ Learn their preferences
- ▶ Payoff function relative to a preference  $u$ :

$$V = u - C + T$$

- ▶ Observed payoff

$$U_t = X_t - C + T$$

# Properties

## Stability

If there are no  $i, j$  such that

$$V(i, M(i)) < V(i, j), \quad V(j, M(j)) < V(j, i).$$

## Fair regret

If for all agents

$$\sum_{t=1}^T \mathbb{E}[U_t(\bar{M}_t(i)) - U_t] \leq O(\ln(T))$$

where  $\bar{M}_t$  denotes the best possible match for  $i$ .

## Social welfare

$$W_t(M) = \sum_i V(i, M), \quad W_t(M) \geq \kappa \max_{M'} W_t(M')$$

# Results

- ▶ Without costs and transfers, fairness-regret is not guaranteed.
- ▶ With costs and transfers, we can guarantee it

## Examples

### Proportional-cost, no transfer

$$C(i,j) = \gamma v(i,j), \quad T = 0$$

The matching is stable.

### Balanced transfer

$$C(i,j) = 0, \quad T(i,j) = \frac{1}{2}(v(j,i) - v(i,j))$$

This is sufficient for all properties, if we use the GS algorithm for stable matching.

### Pricing

$$C(i,j) = c_{i,j}, \quad T(i,j) = g(i), \quad T(j,i) = -g(j)$$

With GS, we guarantee stability and fairness-regret.

# Questions

- ▶ Low optimistic regret seems to imply fairness. . .
- ▶ If we doing GS on the  $i$  side, then only their welfare is maximised. Why not use a different matching algorithm to guarantee high social welfare?
- ▶ All definitions are with respect to the reported preferences. Are any results possible for their latent preferences?